

School fixed effects

SUMMARY

This study analyses the link between student test scores and the school students attend, the policies and practices of the schools, students' family background and their parents' involvement in their education using data from the 2009 wave of the Program for International Student Assessment. We find that (1) a substantial proportion of the variation of test scores within countries is associated with the school students attend; (2) a sizeable proportion of the school fixed effects is associated with school policies and teaching practices beyond national policies or other mechanisms that sort students of differing abilities among schools; (3) school fixed effects are a major pathway for the link between family background and test scores. The implication is that what schools do is important in the level and dispersion of test scores, suggesting the value of further analysis of what goes on in schools to pin down causal links between policies and practices and test score outcomes.

— Richard B. Freeman and Martina Viarengo

School and family effects on educational outcomes across countries

Richard B. Freeman and Martina Viarengo

Harvard University and NBER, The Graduate Institute, Geneva and Harvard's Center for International Development

1. INTRODUCTION

This study uses the Organization for Economic Cooperation and Development's (OECD) Programme for International Student Assessment data set (PISA) to assess the link between student test scores and the school students attend, school policies and practices, students' family background and parental involvement in education. PISA provides data on test scores, schools and family background for hundreds of thousands of students around the world, which makes it the largest cross-country data set in the world for analysing the relation between test scores and their potential determinants. Our focus on schools and school level policies/practices and parental involvement in education provides a more granular picture of the factors associated with student

This is a revised version of a paper prepared for the 58th Panel Meeting of *Economic Policy*, hosted by the Bank of Lithuania. The authors are grateful to the editors, members of the *Economic Policy* panel, the two discussants Daniele Checchi and Marco Manacorda, and three anonymous referees for helpful comments and suggestions. They also thank Ludger Woessmann for sharing his dataset on school tracking. Viarengo gratefully acknowledges support received from the British Academy and the Royal Society.

The Managing Editor in charge of this paper was Tullio Jappelli.

achievement than is found in studies focused on country-level policies such as national standards and external exit examinations.¹

We find that school ‘fixed effects’ are the most important factor associated with student test scores within countries and a major mediator of the relation between family background and scores. To paraphrase the famous campaign slogan of President Clinton, the data tell us ‘it’s the schools, stupid’.² We identify school level teaching practices as reported by students as major correlates of the relation between the schools students attend and test scores, which makes them a potential target for policies to improve outcomes.

There are two pathways by which test score outcomes will vary with the school students attend. The first pathway is through sorting of students of similar ability into the same schools due to national policies on tracking, admissions policies of schools or parent/student or teacher behaviour. Sorting creates differences in outcomes among schools without necessarily affecting the performance of students. The second pathway is through school level educational policies/practices that affect student achievement, with ‘good schools’ raising the test scores of students more than ‘poor schools’.

Differentiating between sorting and educational effects is difficult in international test data. To get some notion of the effect of tracking, we contrast the magnitude of school effects in countries that assign students at early ages to schools beyond their first school to the magnitude of school effects in countries that assign them later. We find evidence of a sorting effect: school fixed effects are larger in countries that assign students to schools earlier than in other schools. But we also find that school effects are large in the countries without such tracking practices, which suggests that the school effect on test performance goes beyond national sorting policies and involves some actual difference in educational policies/practices among schools. We identify some of the relevant policies and practices in the PISA data on the basis of the responses of principals and students to questions about how the school operated.

In the PISA data set, as in virtually all other test score data, students from more advantaged backgrounds perform better on tests than other students. We find that addition of school dummy variables to test score equations substantially reduces the estimated impact of family background on scores. The effects are large in countries with tracking policies and in countries without tracking policies, suggesting that

¹ Woessmann (2003a) and Schuetz *et al.* (2008), find that standards and exit exams affect national test scores. Studies using other data also show that standards and exit examinations are associated with better student performance (Bishop, 1997, 2006; Woessmann, 2003b for the TIMSS 1995; Woessmann, 2005a; Jürges *et al.*, 2005; Greene and Winters, 2005). Studies also find that having an effective link between assessment and curriculum (Bishop, 2006) and keeping high standards (Peterson and Hess, 2006, 2008) improve outcomes, and that increased school autonomy positively affects student performance when combined with an accountability system and external exams (Woessmann, 2005a). An increasing number of countries and states in the US require following a specified curriculum and passing an exit examination for high school graduation (Ou, 2010; Peterson and Hess, 2008).

² http://en.wikipedia.org/wiki/It's_the_economy,_stupid.

schools are major pathway by which parental background relates to student test scores irrespective of the way students are sorted among schools. PISA measures of parental involvement in student education, by contrast, do not account for much of the family background effect on test scores.

The body of the paper consists of five sections. Section 2 reviews the education studies on which our analysis builds. Section 3 describes the PISA data, the equation we use to relate test scores to school and family background variables, and the way we draw inferences about school effects from the data.³ Section 4 presents the estimates of school fixed effects and the impact of school policies and practices on those effects. Section 5 gives our analysis of the role of schools and specific parenting practices as mediators for the link between family background and test scores. Section 6 draws implications for policy.

2. PREVIOUS RESEARCH

This section reviews studies on ‘tracking and other sorting mechanisms’, ‘school policies and practices’, and ‘family background and parental involvement’ that provide the background research on which we build our analysis.

2.1. Tracking and other sorting mechanisms

Researchers in education have long recognized the potential importance of tracking, defined as the explicit streaming of students into different types of schools according to their academic performance, on educational outcomes and the transmission of parental socio-economic position to their children (Betts, 2011). Studies of the effects of sorting on test scores across countries indicate that early tracking increases inequality in educational outcomes without improving national scores (Hanushek and Woessmann, 2006); adds to the impact of family background on educational achievement (Schuetz *et al.*, 2008; Woessmann *et al.*, 2009) and labour market earnings (Brunello and Checchi, 2007). But national tracking policies are not the only source of sorting students among schools. Parents/students can self-sort into particular schools, from private schools to public schools of differing quality; teachers of different qualities can be recruited and choose to work in particular schools; and the government or private groups can allocate other educational resources differentially among schools.

With respect to student self-sorting, Nechyba (2011) summarizes the following mechanisms: parental background leading to residential segregation, parental choices of private as opposed to public schools and of schools of different quality in the sectors. Nechyba notes that ‘while sorting mechanisms may differ across countries, there is

³ Woessmann (2005b) compared the importance of family background of five Asian countries (Hong Kong, Japan, Singapore, Korea and Thailand) to three Western countries (France, Spain and the US) in the TIMSS 1995 study.

little reason to doubt that sorting itself is a critical component of most primary and secondary school systems' (p. 1029).

To the extent that residential areas correlate with the quality of schools in the area in which a family lives, which will depend on family income and the housing market, this can create substantial sorting (Bjorklund and Salvanes, 2011). Housing markets also play a similar role in cases where admission does not depend on the residential area but transportation is costly (Nechyba, 2011). Another family-based mechanism for sorting is through parental education. More educated parents presumably have better information on the different quality of schools (Bjorklund and Salvanes, 2011, p. 215), and will push harder for getting their children into the better schools compared to less educated parents. Sorting related to education may be closely related to sorting through economic resources but may also be independent of income as suggested by Bjorklund and Salvanes (2011). And the ability or other characteristics of students will likely play some role in these modes of sorting.

Finally, there is the sorting between public and private schools, which researchers have explored both with observational and experimental methods (Epple and Romano, 1998; Angrist *et al.*, 2002; Altonji *et al.*, 2005). This source of sorting may be related to family background and income, student ability or other characteristics. To the extent that the private school sector is well funded, it may work as a sorting mechanism even for young persons of modest means, by giving scholarships to the more able, and even for young persons of modest academic ability, by giving scholarships for athletic or other talent.

Students aside, the decisions of schools to offer teaching jobs to teachers of differing characteristics and teachers' acceptance or rejection of job offers can create sorting of teachers across schools that will impact student performance. Hanushek and Rivkin's (2006) review of studies on teacher labour markets acknowledges the importance of measuring teacher quality and the difficulty related to selection and unobserved characteristics: 'Credible research into training versus selection issues as related to certification policies, merit pay, and so forth clearly requires longitudinal observations that link teachers, programs, and student performance. Until recently, there has been little possibility of such work' (p. 1073). A similar problem is found in the estimation of teachers' earnings equations (Dolton, 2006). Still, using a data-rich data set on teachers in New York, Lankford *et al.* (2002) found that disadvantaged students are disproportionately in classrooms with less skilled teachers, especially in urban areas. They relate this to district differences in preferences for teaching qualifications and hiring efficiency (which will depend on the school budget), on the demand side, and teacher preferences on the supply side. They note, however, that 'we know very little about sorting or the causal relationships that lead to sorting' in these situations (Lankford *et al.*, 2002, p. 39).

With respect to the allocation of other resources, schools with high performing students may obtain more and better resources while gifted students may benefit from special programmes within a school or admission to schools with special resources –

for instance, a special maths or science academy or school for music and performing arts. But the allocation of resources may be of a compensatory nature (e.g. Holmlund *et al.*, 2010; West and Woessmann, 2006). In this case, more resources are allocated to more disadvantaged and low-performing students.

2.2. School policies and practices

The educational literature on the relation of school policies and practices to student performance provides a benchmark for our findings.

Studies of school selectivity invariably find that more selective schools have higher performing students than less selective schools and lower variation in scores among students (Sprietsma, 2008; Jackson, 2010; Pop-Eleches and Urquiola, 2013). Studies of whether sorting students by classrooms within the same school benefits students by providing more targeted instruction by ability or harms students by reducing the interaction between low performing and high-performing students give inconclusive results (Betts and Shkolnik, 2000; Figlio and Page, 2002).

Studies of *School Autonomy* and *Accountability*⁴ suggest that the alignment of autonomy and accountability leads to higher outcomes (Fuchs and Woessmann, 2007, Woessmann *et al.*, 2009) but have an unclear impact on inequality of outcomes among students (Hanushek and Raymond, 2003). Hanushek *et al.* (2013) shows that for 42 countries that participated in PISA from 2000 to 2009, school autonomy in academic content, hiring and salaries is associated with higher student learning outcomes in developed countries.

Studies of *Resource* inputs such as spending per student, teacher-student ratio, class size and instructional time have produced controversy. Reviewing extant studies through the early 2000s, Hanushek (2006) concludes that in advanced countries there is no compelling evidence that changes in the level of any of these resources, within the observable range of variation, have a substantive effect on educational outcomes. The Tennessee STAR study, however, gives a different picture – with a one-third reduction in class size associated with higher student achievement by the equivalent of three additional months of schooling (Krueger, 1999). But other experiments find smaller or no impacts of class size (Chingos and Whitehurst, 2011; Altinok and Kingdon, 2012; Woessmann, 2010). Studies relating to teachers' characteristics such as experience, academic and professional qualifications, find that while teachers matter in the sense that individual teacher fixed effects are important in explaining student test scores, observed characteristics are generally unrelated to teachers' performance (Hanushek and Rivkin, 2006).

⁴ Research in this area using the international studies has examined various dimensions of school autonomy (e.g., autonomy over resources allocation, hiring and firing decisions), regulations of school governance and standardized exams (e.g., periodic measurement of teacher and principal performance, standardized exams).

With respect to school *Practices*, Bratti *et al.* (2011) find that cooperative behaviour among students through group learning improves educational outcomes. By relying on a quasi-experimental approach for students in primary and middle schools in Israel, Lavy (2011) finds a positive impact on learning outcomes associated with teaching practices which include group learning activities and activities that focus on developing students' analytical and critical skills. Examining learning practices in several waves of TIMSS, and PIRLS, Algan *et al.* (2013) find that 'horizontal teaching practices' (i.e., where students work in groups, ask questions) is related to pro-social beliefs while 'vertical teaching practices' have the opposite effect. Comparing student reports of teaching practices Bietenbeck (2011) reports that memorization and lecture-style teaching produce better outcomes whereas selected teaching practices such as working in groups, students' autonomy in problem-solving decisions have a small insignificant effect on learning outcomes. Schwerdt and Wuppermann (2011) compare lecture-style teaching to problem solving and find a positive association between the practice of lecture-style teaching and students' test scores.

2.3. Family background and parental involvement

Studies of the educational achievement of young persons invariably find that children from more advantaged backgrounds perform better on average in school in terms of test scores and other school outcomes, albeit with heterogeneity in the magnitude of the impact of background factors among countries (Freeman *et al.*, 2011; Hanushek and Woessmann, 2011a) and with differences related to the measure of background. Ammermuller *et al.* (2005) and Woessmann (2005b, 2008) use parental education to analyse background effects on children's education. Other researchers favour the books at home measure as reflecting not only the parents' scholarly culture but also the socio-economic background of the household (Di Maggio, 1982; Dronkers, 1992). Ammermuller *et al.* (2005), Woessmann (2003a, 2005b, 2008) and Schuetz *et al.* (2008) use books at home in analyses of educational outcomes in different waves of TIMSS while Peterson and Woessmann (2007) analyse it in the PISA 2003. Other scholars have relied on the socio-economic and cultural status index. Jenkins *et al.* (2008) and Machin and McNally (2012) among others report significant relation between this index and learning outcomes.

Research on parental participation in home and school-related activities finds a positive association between parental involvement and students' learning outcomes (Bourdieu, 1986; Coleman, 1988; Hoover-Dempsey and Sandler, 1997). But Desfor- ges and Abouchaar's (2003) note the limitation of evaluations which do not satisfactorily address the endogeneity of parental involvement. Parents with greater education and resources are likely to invest more time and resources in their children's education than less educated parents or those with fewer resources. Parents are also more likely to invest in the education of children who are more receptive and/or in children having problems in school – for instance, meeting with teachers or providing help with

homework. Country-specific longitudinal studies (e.g., the UK National Child Development Study, the British Cohort Study, US National Educational Longitudinal Study) suggest that parental participation in learning activities at home (Schoon and Parsons, 2002), aspirations for children's continuing in education (Singh *et al.*, 1995), discussion of school experience and involvement in cultural and learning activities (George and Kaplan, 1998) are positively associated with children's educational performance conditional on the socio-economic status of parents.

3. PISA DATASET AND EMPIRICAL ANALYSIS

3.1. Data

Beginning in 2000 the PISA study has tested 15-year-old students on their skills, knowledge and ability to use this knowledge to solve problems in real-life situations every three years.⁵ The tests cover mathematics, science and reading. Participating countries randomly select schools to administer the test and randomly select students within the schools to provide a nationally representative sample of students. PISA standardizes test scores to a mean of 500 and standard deviation of 100 on the basis of the 2003 PISA. Differences in test scores can thus be interpreted in terms of percentage points of an international standard deviation. An increasing number of countries have participated in each succeeding wave, including developing countries outside the OECD.

We use data on students and schools from the 2009 wave. This data set provides information on about 520,000 students in 18,641 schools from 74 economies,⁶ of whom 34 are OECD countries and 31 are 'partner countries' outside the OECD. In addition to the PISA tests, students fill in a questionnaire on their characteristics, family structure and background; while principals of each participating school report on school characteristics, policies and practices. In addition, 14 countries administered a questionnaire to parents regarding their involvement at home with the child and in

⁵ Students' age varies between 15 years 3 months and 16 years 2 months and to be eligible to participate must have completed at least six years of formal schooling (OECD, 2010a).

⁶ Around 470,000 students in 65 countries participated in PISA 2009. In addition to this, about 50,000 students were tested in 2010 in the framework of PISA 2009+ (from 7 countries and 2 Indian states (Himachal Pradesh and Tamil Nadu). This was the same assessment as PISA 2009 but carried out one year later. More information on PISA 2009 and PISA 2009+ is available at <http://mypisa.acer.edu.au/>. This is the list of countries and economies: Albania, Argentina, Australia, Austria, Azerbaijan, Belgium, Brazil, Bulgaria, Canada, Chile, Chinese Taipei, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Great Britain, Greece, Himal Pradesh – India, Hong Kong, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kazakhstan, Korea, Kyrgyzstan, Latvia, Liechtenstein, Lithuania, Luxembourg, Macao-China, Malaysia, Malta, Mauritius, Mexico, Miranda-Venezuela, Moldova, Montenegro, Netherlands, New Zealand, Norway, Panama, Peru, Poland, Portugal, Qatar, Romania, Russian Federation, Serbia, Shanghai-China, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Tamil Nadu – India, Thailand, Trinidad and Tobago, Tunisia, Turkey, United Arab Emirates, United States, Uruguay.

school-related activities,⁷ including involvement when children were in primary schooling (ISCED 1).⁸

We analysed test outcomes in mathematics, reading, science and some non-cognitive measures of performance.⁹ The results for the different test scores are sufficiently similar that we report in this paper only those for mathematics. Results for the other tests are available on request from the corresponding author.

We analyse students in the full set of countries in the PISA and then group the countries into five subsets: OECD economies, European Union economies, Asian economies, other high income economies, and middle income economies.¹⁰ Countries can appear in more than one subset in our groups: for instance the EU countries are also part of the OECD country subset. We differentiate the Asian countries because five of the top ten scoring countries in PISA 2009 were Asian, which raises questions about how they differ from other countries in educational practices.¹¹

3.2. Empirical framework

At the heart of our analysis are estimates of variants of the following equation:

$$T_{isc} = \alpha_1 F_{isc} + \alpha_2 FP_{isc} + \beta_1 \mathbf{X}_{isc} + \beta_2 \mathbf{S}_{sc} + \beta_3 \mathbf{SD}_{sc} + \mu_c + \varepsilon_{isc} \quad (1)$$

where T is the test score of student i in school s of country c ; F is a measure of family background characteristics for student i in school s of country c ; FP is a measure of the parenting practices for student i in school s of country c ; \mathbf{X} is a vector of the student's characteristics, for student i in school s of country c ; \mathbf{S} is a vector of the policies and practices of school s of country c ; \mathbf{SD} is a vector of dummy variables for each school in

⁷ The data set is publicly available from the PISA website: <http://pisa2009.acer.edu.au/>.

⁸ Throughout, we use the sampling weights in the database as described in PISA Technical Report 2009 (OECD, 2012a). To deal with missing observations, we kept all students with complete test score data and imputed the values of missing explanatory variables by using mean values and including in the regression an indicator for missing data (see Hanushek *et al.*, 2013).

⁹ Our analysis of non-cognitive learning outcomes considered indicators related to disruptive behaviour and two psychological traits central to the five factors model (Costa and McCrae, 1992; Peterson and Viarengo, 2011): openness to learning (attitude towards school) and locus of control (self-confidence in ICT high-level tasks). Non-cognitive skills have been largely neglected in the cross-country studies due to the paucity of information available and difficulties of measurement (Woessmann *et al.*, 2007; Brunello and Schlotter, 2011).

¹⁰ Footnote 6 provides the list of countries in our analysis. Korea and Japan are Asian countries members of the OECD. Mexico is a middle-income country member of the OECD. For the purpose of the analysis that follows Korea and Japan are included in the Asian group and Mexico in the middle-income group.

¹¹ The Asian countries have large shares of students at the proficient level 5 (OECD, 2010b). PISA classifies students into five proficiency levels, 'moderate' performers achieve proficiency levels of 2 or 3 whereas 'top' performers achieve proficiency levels of 5 or 6. OECD (2010a, p. 8; Peterson *et al.*, 2011, p. 8). More than 20% of students in Korea, Japan and Hong Kong perform above proficiency level 5 (OECD, 2010c, p. 63) and above the advanced international benchmark of 625 or better. Cross-national surveys of learning achievement show that Korea, Macao, Shanghai combine top scores with levels of inequality below the OECD average (e.g., OECD, 2010b, p. 53).

the data set;¹² μ_c is a vector of country dummy variables; ε is an error term;¹³ and α_1 , α_2 , β_1 , β_2 , β_3 are vectors of coefficients to be estimated.

Because PISA gathers data on test scores from about 35 students in each school, we have two ways to examine the relation between the schools students attend and test scores. The first way is a ‘sibling’ type estimate based on estimating the school level dummies **SD**. To obtain the fixed school effect we omit the policy and practice vector in Equation (1) (that is, we set β_2 at 0). Just as sibling studies of family background effects use the fact that siblings grow up in the same family to estimate the variance in outcomes associated with family background without measuring that background,¹⁴ the PISA data allow us to estimate school effects through the similarity of outcomes among students in the same school absent measures of school policies and practices. The second way to examine the effects of schools on outcomes is to estimate the coefficients in the **S** vector of policies and practices in Equation (1). To do this we omit the school dummy variables (that is, set β_3 at 0). The variance in test scores due to measured policies/practices will invariably fall short of the variance in test scores in school fixed effects, since the measured variables do not account for all of the differences among schools.

If students were randomly assigned to schools, the difference between the contribution to the variance of the estimated school fixed effect and the contribution to the variance in scores associated with observable school practices and policies would reflect the effect of schooling on outcomes through mechanisms rather than observable policies and practices that affect student education. But, as noted in our discussion of tracking, the assignment of students among schools is not random. Students of similar ability are likely to be sorted among schools, producing school fixed effects independent of what schools actually do for students.¹⁵

To help differentiate the variance in test scores among schools due to sorting from the variance to what schools do to students, we first examine the effect of different measures of sorting on variances and then examine the effect of explicit policies on the variance of schools under different sorting regimes.

Approaching the problem from the sorting perspective, we contrast differences in the school fixed effects on test outcomes between countries with early sorting policies and countries with late sorting policies. We expect greater variation in the average

¹² In the estimation of the school fixed effect models country fixed effects are absorbed by school fixed effects.

¹³ Clustering-robust linear regression (CRLR) has been used in the estimation to estimate standard errors given the fact that the error term has both school-level and student-level components.

¹⁴ Bjorklund and Salvanes (2010) report that factors shared by siblings explain more than 50% of the overall variation in years of schooling for children in a sample of developed countries.

¹⁵ If different schools had no discernible effect on student performance, parental efforts to get their students into better schools would make little sense in terms of their education, so sorting practices by parents are presumably indicative of actual educational effects. Demonstrably better school effects on students should induce parents to make greater efforts to get able students into higher quality schools. We do not explore this interaction in this paper.

scores among schools in countries with early tracking policies than among schools in countries without those policies. Following a similar logic, we also contrast variance in test scores among schools with different admission policies and openness to residential and parental choice to see whether those measures of sorting account for the bulk of school fixed effects.

Approaching the problem from the perspective of how schools affect student performance, we consider the contribution of school fixed effects and measures of explicit policies on test scores in countries/schools with different sorting policies. Large school fixed effects in countries or among schools *without* tracking policies would indicate that something other than tracking underlies school differences. But even more telling would be evidence that explicit policies and practices affect outcomes similarly among countries with and without tracking. It is difficult to argue that tracking is the whole story for school fixed effects if the evidence shows sizeable effects absent tracking and relations between test scores and school policies or practices under different sorting regimes.

4. SCHOOLS EFFECTS: SORTING VERSUS SCHOOL POLICIES AND PRACTICES

4.1. The role of schools

Table 1 records statistical properties of the maths test score on which we focus. Column 1 records the mean and standard deviation (in brackets below the mean) of the test score in all the countries in the data set and in countries in the five country groupings. The mean scores for the poorest countries in our sample, ‘middle income countries’ by the World Bank definition, fall below those of the other and higher income countries. The mean scores for the Asian countries are higher than the scores for other countries. The coefficient of variation of the test scores in column 2 shows modestly higher variation in scores among the middle income countries than others and modestly lower variation in test scores for the Asian countries than for the others. Variation of scores is, however, large among all groups, which reflects the wide distribution of scores among students within countries.

The columns under the heading ‘Percentage of Variance in Scores’ record the proportion of the variance in test scores associated with country, measures of background, and schools.¹⁶ The proportions of variance are calculated separately for each of these factors. The column labelled country shows the percentage of variance from regressions of student test scores on country dummies. The column labelled background shows the percentage of variance from regressions of test scores on the

¹⁶ Standard errors in the empirical analysis here and in the remainder of the study are corrected for clustering.

Table 1. Test scores and analysis of variance decomposition of test scores: country, family background and school effects, PISA 2009

| Country groups ^g (number of countries and students in parentheses) | Mean and variation in scores ^a | | Percentage of variance in scores due to specified factors ^c | | |
|---|--|--|---|--|---------------------------------------|
| | Average test score (standard deviation) | Coefficient of variation ^b | Country effects ^d (C) | Background effects ^e (B) | School effects ^f (S) |
| All ^h (74; 515,956) | 449.47 (104.74) | 23.30 | 0.322 | 0.334 | 0.618 |
| OECD (31; 249,125) | 489.30 (94.54) | 19.32 | 0.053 | 0.276 | 0.435 |
| EU 15 ⁱ (17; 150,696) | 498.24 (95.23) | 19.11 | 0.024 | 0.291 | 0.475 |
| Asian (7; 38,095) | 540.28 (96.02) | 17.77 | 0.026 | 0.143 | 0.502 |
| Other high income (8; 44,384) | 465.71 (86.47) | 18.57 | 0.021 | 0.191 | 0.348 |
| Middle income (27; 179,366) | 388.28 (81.22) | 20.92 | 0.076 | 0.258 | 0.522 |

Note: Sampling probability weights have been used in order to provide descriptive statistics representative for each country

^aTest scores in mathematics here and in the remaining of the analysis unless otherwise specified.

^bCoefficient of variation (CoV) is a normalized measure of dispersion calculated as the standard deviation divided by the mean score.

^cVariance has been extracted as the share of test score variation explained by the model (ANOVA/R²). In the underlying regression test score is a function of country dummies. The figure has been divided by test score variance.

^dIn the underlying regression test score is a function of country dummies. The figure has been divided by test score variance.

^eIn the underlying regression test score is a function of the family background measure, student-level and school-level controls. Here family background has been proxied by the ESCS index. The figure has been divided by test score variance. The ANOVA analysis has also been carried out by using the components of the ESCS unrestrictedly. Results are consistent with the analysis presented here and available from the authors.

^fIn the underlying regression test score is a function of school dummies. The figure has been divided by test score variance.

^gOnly a low-income country participated in the PISA 2009 study: Kyrgyzstan, so we did not include this group. The analysis has been carried out for this country and is available from the authors upon request.

^hThis group includes 72 countries and 2 Indian states.

ⁱThis category includes Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom, Norway and Switzerland.

family background measures listed in the table note. The column labelled schools shows the percentage of variance from regressions of test scores on school dummies.

Country contributes substantially to the variance in the whole data set because the middle income countries have much lower test scores than the other countries. But within each of the groups country is associated with only a small proportion of the variance in scores. Background factors have a substantial impact on the variance in the overall sample and within every category but contribute least to the variance among the Asian countries. The largest source of variation in test scores in Table 1,

however, is *the school the student attends*.¹⁷ The regression of test scores on the school dummies explains a large proportion of the variation of scores in all groups. Estimates of the contribution of schooling to the variance of test scores separately by country confirm this result for all countries, albeit with variation in the magnitudes.¹⁸ That the students' school is the biggest measurable contributor to the variance of test scores motivates ensuing analysis.

The first question we ask about this empirical result is the extent to which school fixed effects result from national tracking policies or other sorting practices as opposed to the potential impacts of schools on learning by students. If the variance in test scores among schools were due largely to sorting of students by prior ability/knowledge, the school effects would arguably be more indicative of efforts to preserve existing social connections and thus provide little insight into educational policy. If, by contrast, a sizeable part of the school fixed effects is due to school policies and practices, the finding directs attention at the impacts of measured policies and practices and ways to transfer best practices to more schools, and to further investigation of policies and practices that the PISA does not measure.

As a first step toward measuring the extent of sorting versus educational policies/practices on the school fixed effects result we built on the research referenced in Section 2 to estimate the effect of national policies toward tracking of students into higher level schools on the variance of test scores. We gathered data on the first age at which students move from one level of schooling to the next level in different countries.¹⁹ Researchers who study national tracking systems use these data to divide countries into two groups: countries which sort students at ages less than 16 are viewed as having tracking systems while countries with a first age of sorting above age 16 are viewed as not having a tracking system.²⁰ Given that the students in the PISA are 15–16 years old, this definition works well with our data. But rather than settling for a dichotomous division, we created a more refined measure of tracking by differentiating countries by the first age of movement to a different school. If tracking policies sort students by abilities and these measures reflect those policies, we expect that the variance of test scores *among schools* will be higher in countries that tracked students early in their lives than in countries that tracked them later in their lives. By contrast, the *within-school* variance of test scores should be lower in the tracking countries than in the non-tracking countries.

¹⁷ This can be explained either by the effect that schools have on determining student's learning outcomes or by some sort of self-selection that affects the student composition into the different schools. In the analysis presented in the following sections of the paper different possible explanations are going to be investigated.

¹⁸ These estimates are available from the authors.

¹⁹ First age at which selection takes place in the education system (OECD, 2005, Table D6.1; Hanushek and Woessmann, 2006; Brunello and Checchi, 2007; Woessmann, 2005b; in addition to this, Professor Ludger Woessmann kindly shared with us his international dataset on tracking acquired through detailed personal inquiries). This information refers to the policy set at the national level.

²⁰ For a review of the existing studies see Hanushek and Woessmann (2011a, section 4.4.4).

Accordingly we examined differences in the relation between test scores and schools at the level of countries and at schools within countries. At the country level, we calculated the coefficients on the vector of school dummies from Equation (1), absent measures of school practices and policies. The estimated coefficients measure the fixed effects impact of each school. This is equivalent to taking the average performance of a school adjusted for the effects of student background factors. Then we related the estimated school effects to the rank of the school in the country's distribution of school scores from highest to lowest. The basic notion is that tracking should produce a steeper relation between the ranking of schools in the country's distribution of schools by test scores and the estimated school effects.

The regression of the estimated school fixed effects on the rank of the school for each country yields a coefficient α_1 for country i that shows how much being higher in the distribution of schools affects the test score:²¹

$$\beta'_{3sc} = \alpha_1 R_{sc} + \varepsilon_{sc} \quad (2)$$

where β'_{3sc} is the vector of estimated fixed school effects for school s in country c (i.e., the mean test score in the school adjusted for the other variables in regression (1)) and R_{sc} is the rank of school in the distribution of test scores for the school s in country c .

Because the PISA sample includes several countries with huge numbers of schools – Mexico (1,535 schools), Italy (1,097 schools), Brazil (947 schools) and Canada (978 schools) – the range of ranks for those countries diverges hugely from the range of ranks for the typical PISA country that has in the order of 100–200 schools. The result is that estimates of Equation 2 for the countries with huge numbers of schools give much smaller estimated α_1 coefficients than for the other countries. To deal with this scaling issue and standardize the data for cross-country analysis, we created a new data set consisting of the estimated fixed effect for each decile of schools for each country.²² This gives us 10 average test scores for each decile and the ranks from 1 to 10 for the deciles for each country. We then regressed the score of the 1st decile, 2nd decile, ..., 10th decile school on the ranks of their corresponding decile. This gives us a slope coefficient for each country estimated on a similar number of observations and with a similar range of the right-hand side rank variable.

²¹ Summarizing the relation between test score and school ranking with a single regression parameter in this manner forces a simple linear relation onto the rank-score curve. The actual curves show more complex logistic patterns, which depend on the nature of sorting, groups sorted, the number of students in schools and the bounds on school scores (which lie between the lowest and highest PISA scores multiplied by the number of students in bottom and top schools). But for the purposes of examining whether or not formal tracking systems steepen the rank score relation, the linear analysis suffices.

²² We thank the discussants for their suggestion on how to deal with this problem.

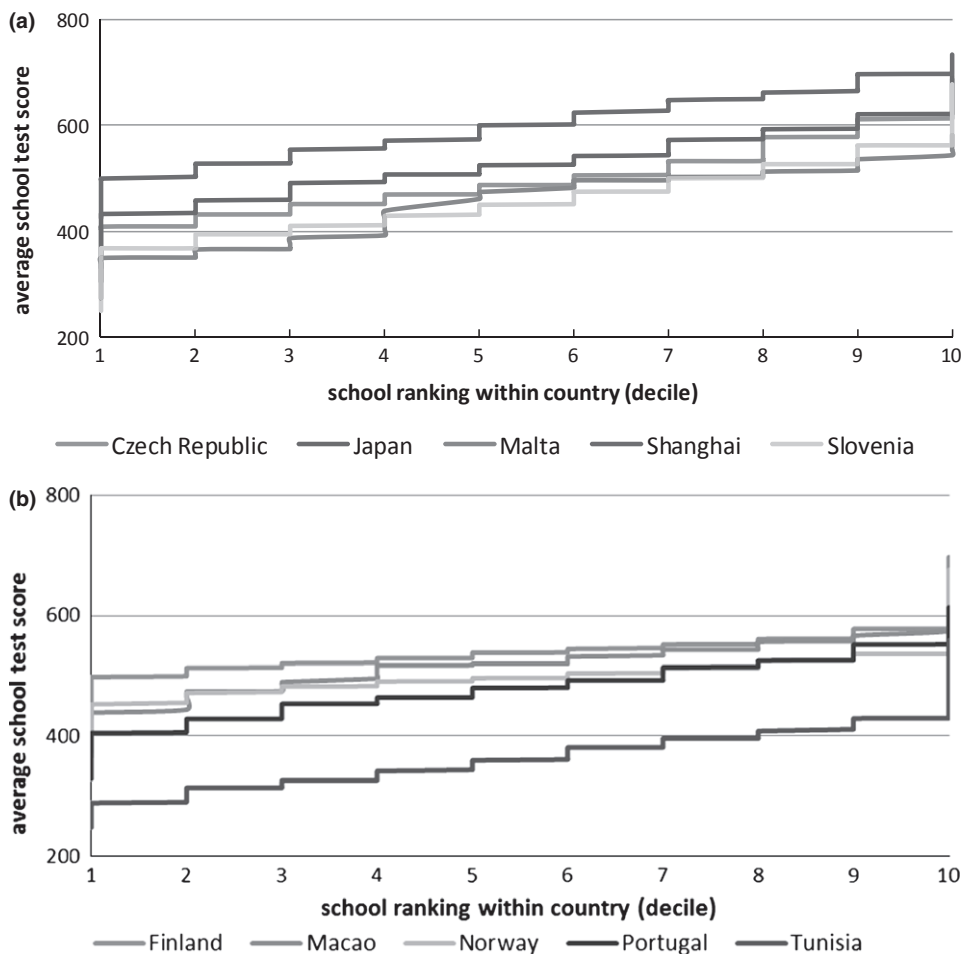


Figure 1. School rankings and average test score, PISA 2009 (a) Top 5 – countries with largest estimated coefficient on school rank; (b) Bottom 5 – countries with smallest estimated coefficient on school rank

Notes: The picture at the top (bottom) shows the school ranking and the related average school test score for the five countries with the largest (smallest) estimated coefficient on school rank. The graph for all the 74 countries and economies in our sample – including Costa Rica, for which the data on tracking is not available – is available from the authors.

Figure 1 graphs the scores of schools against the ranks of schools in the curves for the five countries with the largest gradients and the five countries with the smallest gradients based on the decile regressions.²³ By construction the curve rises monotonically, but the gradients of the rank-score curve vary substantially among countries. Countries with the largest gradient have an early tracking system whereas countries

²³ The estimated coefficient on school rank for each country in our sample is available from the authors. The coefficients have been extracted from the country-specific regressions run at school-level. The graph for the remaining countries and territories in our sample is available from the authors.

with the lowest gradient have late tracking systems, Portugal being a noticeable exception.²⁴ The estimated slopes for countries are available from the authors and show high gradients for countries of Continental Europe (France, Germany, Hungary, the Netherlands), heterogeneous gradients for Asian countries and smaller gradients for Anglo-Saxon (Australia, Canada, Ireland) and Nordic European countries (Denmark, Finland, Norway).

To check the robustness of the country ordering by the gradient of the rank-score figure, we estimated rank-score curves for countries in three other data sets in periods close to the PISA 2009: TIMSS 2011 4th grade mathematics students (41 countries overlap), TIMSS 2011 8th grade mathematics students (30 countries overlap), and PIRLS 2011, 4th grade reading (19 countries overlap).²⁵ Differences among the studies in design and structure as well as grade groups, notwithstanding the ordering of countries by the estimated rank-score gradient is consistent with that in PISA, with the greatest similarity between PISA and the TIMSS 4th grade, TIMSS 8th grade test and the least with the PIRLS 4th grade test.

By the logic of tracking policies, countries that track students earlier should have steeper gradients linking school fixed effects to the rank of schools in the distribution of scores than countries. Similarly, the proportion of the variance of student test scores due to school effects as opposed to within school variance should be higher in countries which do more tracking.

The regressions in Panel A in Table 2 test this expectation with two measures of country tracking policies. The first measure is the standard measure of tracking systems in the cross-country literature: a dummy variable for whether the first age at which students move school levels is below 16. The second measure is the 'tracking index' based on the age that countries begin tracking is available from the authors. The ages vary from 10 to 18, but our index reverse codes them to be consistent with the dummy variable measure so that higher numbers reflect stronger tracking (i.e., the index is 18 for schools that send students to another school level at age 10, 17 for those that do so at age 11, and so on). The Appendix provides a description of these tracking variables used for the analysis.

The first two columns of Panel A of Table 2 give the results of regressing the estimated gradients of school test scores on school rank on the specified measures of tracking. The first regression excludes covariates for student characteristics. The second regression includes country averages of the covariates listed in the table note.²⁶ The positive coefficients on the tracking dummy variable and on the tracking index show that earlier sorting of students into another level of schooling is associated with a

²⁴ Age at first selection in Portugal is equal to 15.

²⁵ TIMSS refers to the Trends in International Mathematics and Science Study whereas PIRLS refers to the Progress in International Reading Literacy Study. Both studies are administered by the International Study Center, Lynch School of Education, Boston College, and the International Association for the Evaluation of Educational Achievement. More information is available at <http://timss.bc.edu/>.

²⁶ Descriptive statistics of these covariates are presented in Tables A1 and 6.

larger gradient of test scores on school rank. A high ranking school has a higher test score compared to others in the same country when the country tracks earlier relative to a high ranking school in a country which tracks students later.

The next two columns give the results of regressing the ratio of the variance in test scores attributed to the school dummies over total variance in each country on the two measures of tracking policies with and without covariates for country characteristics of students. These regressions show that the proportion of variance between schools is larger among countries with lower ages of sorting.

Panel B of the table shifts from country level regressions to school level regressions to assess the effects of tracking policies on test scores. Here, we regress the estimated school fixed effects on the rank of each school in their own country and on country dummy variables *separately* for countries with early tracking policies and for countries with later tracking policies. This specification forces a single coefficient on school rank for all of the countries in each of the groups. To deal with the different number of schools in this analysis, we have weighted the observations by 1/number of schools in a country. The results confirm the regression findings in the country data. The slope of the test score on the rank of schools is higher in countries with an early tracking policy than in countries without such a policy when the tracking measure is the sole explanatory variable and when we include the covariates listed in the table note. We also regressed the test score on the rank of school, the age-based measure of the tracking policy and the interaction of the rank with the policy measure. If tracking produces a steeper gradient, the estimated coefficient on the interaction term should be positive. As the last line of Table 2 shows, this is the case. Countries with stronger tracking policies sort students among schools in ways that produce greater differentiation in average test scores among schools.

The variance in test scores among schools in countries that do not track students before age 15 cannot be associated with explicit tracking policies. To be sure, not all of the variance in test scores among schools in these countries will be due to true school impacts on students, but the school fixed effects in these countries should be more reflective of true school impacts than the school fixed effects in countries with early tracking policies. Subtracting the maximum estimate of the effect of tracking on the variance of test scores in Table 2 – the 14 point impact in the single variate regression – from the Table 1 estimated impact of school fixed effects on the variance of test scores of 62% would reduce the fixed school share of the variance in test scores among students to 48%. This still leaves schools as the major factor associated with student test scores.

4.2. Other sources of sorting

Per the discussion of sorting earlier, students may sort among schools through mechanisms beyond national tracking policies and thus contribute to school fixed effects. Some countries allow parents greater choice in their enrolment. Schools may have

Table 2. Estimates on effect of country tracking practices on gradient of effect of rank-school school-fixed-effects and percentage of country variance in test scores, by country tracking practices

Panel A: Country level regressions

| Measure of tracking in country ^a | Coefficient from regression of gradient of school test score–school rank on tracking policies Country-level regressions ^b (n = 68) | | Coefficient of percentage of country test score variance attributable to variance in scores across schools on tracking policies (n = 68) | |
|---|---|---|--|---|
| | Single variate regression | Regression with covariates ^c | Single variate regression | Regression with covariates ^c |
| School tracking policy Dummy = 1 for tracking | 4.12*** (0.92) | 2.59*** (0.95) | 14.40*** (3.16) | 7.07*** (2.58) |
| Tracking Index based on age | 1.11*** (0.21) | 0.83*** (0.25) | 3.74*** (0.79) | 2.54*** (0.78) |

Panel B: School level regression

| Country group | Regression coefficient of school-fixed-effects on school rank, with country dummies ^d (number of schools = 17,539) | |
|---|---|---|
| | Single variate regression | Regression with covariates ^e |
| School tracking policy = 0 [no. schools=6,288] | 0.25*** (0.003) | 0.20*** (0.004) |
| School tracking policy = 1 ^f [no. schools=11,251] | 0.36*** (0.005) | 0.25*** (0.005) |
| Rank interacted with tracking index based on Age * Rank [all schools] | 0.009*** (0.002) | 0.002 (0.001) |

Notes: Significance level (based on clustering-robust standard errors): * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

^aAnalysis for 68 countries for which we have tracking information. Basic relations in the data similar to those with regression results for the entire sample. First age at which selection takes place in the education system. This information refers to the policy set at the national level. Several sources have been used to extract this information including Professor Woessmann who kindly shared his dataset with us that was constructed through detailed country-specific inquiries (OECD, 2005, Table D6.1; OECD, 2010d, Table IV 3.2a; Hanushek and Woessmann, 2006; Brunello and Checchi, 2007, Schuetz *et al.*, 2008; Woessmann, 2005b; country-specific sources for Malta and India). Information on tracking was not available for the following countries: Costa Rica, Georgia, Kazakhstan, Mauritius, Miranda – Venezuela, Serbia.

^bThe analysis is carried out at the country-level and the dependent variable is the country-specific coefficient on school rank. These country-specific coefficients are available from the authors.

^cThese are country averages of: student's age, gender, grade, whether the language of test is the main language spoken at home, whether the student lives with two parents (natural or otherwise), immigrant status, parents' immigrant status, total school enrolment, dummies for community size where the school is located, index of socio-economic status (ESCS). Descriptive statistics related to these variables are presented in Tables A1 and 6. Student sampling weights are included.

^dThe dependent variable is school fixed effects. The coefficient presented in the table is related to the school ranking. Country dummies are included. School-level regressions are weighted by 1/number of school.

^eThese are school averages of the control variables described in note c.

^fSchool tracking is equal to 1 if age at first selection is below 16.

selective admission policies, limiting enrolment to high performing students. Parents may affect student placement by residing in areas with reputedly higher quality schools.

The PISA principals' questionnaire contains three questions which allow us to assess the potential impact of these forms of selection on the estimated school fixed effects. The first question relates to the governance of schools. It asks whether the school is public or private. Private schools are defined as 'schools managed directly or indirectly by a non-government organization (e.g. a church, trade union, business, or other private institution)'. The PISA database also provides information on the sources of funding of the school. Following Woessmann (2009) we classify the public and private sources of school funding.²⁷ Assuming that families with more academically able students get their children into schools with stronger academic reputations, which in some countries are private institutions, we expect a positive association between private schools and test scores and, assuming those schools have greater homogeneity of students than others, a smaller coefficient of variation than in other schools.²⁸ The second question, extracted from the school principals' questionnaire, is 'how often is a student's record of academic performance (including placement tests) considered when students are admitted to your school?' Schools that take account of performance in admissions policy should have a more homogeneous student body and correspondingly lower variance of scores within the school than other schools.

The PISA has a third question regarding school policies that has a more ambiguous relation to the variance of school test scores. It asked principals 'How often is residence in a particular area considered when students are admitted to your school?' If residence²⁹ is a strong factor in admission, a school will likely admit nearly all students in its catchment area – the less able as well as the more able, and thus be less selective on academic performance, producing greater within school variance in test scores and smaller variance in school fixed effects. On the other side, if wealthier and more educated families live in different catchments than poorer and less educated families, such a policy would increase the variance in school effects. If families with more academically able students move to areas with better schools, the dependence of residence on schools would add yet another sorting contribution to the variance of

²⁷ Specifically, we classify as public funding the share of funding provided by the 'Government (includes departments, local, regional, state and national)' and as private the funding provided by the remaining categories ('Student fees or school charges paid by parents', 'Benefactors, donations, bequests, sponsorships, parent fund raising' and 'Other').

²⁸ Given differences in the management and funding of private and public schools across countries, this expectation may hold only in part of the sample.

²⁹ In order to further examine the sorting mechanism through residential location the empirical analysis has been carried out for a finer set of countries. These four countries do not have national tracking policies but two are more egalitarian (Finland and Norway) whereas in the other two residential sorting appears to play a more important role (United Kingdom and United States). Results related to this analysis are available from the authors upon request. We also acknowledge the fact that other sorting mechanisms may play an important role (e.g., parental choices related to public vs private schools).

Table 3. Regression coefficients of test scores, school fixed effects and coefficient of variation of scores within schools on sorting mechanisms, PISA 2009^a

| Variables ^b | Share of schools with practice | School test score (1) | School fixed effects ^c (2) | Coefficient of variation of scores within schools (3) |
|-----------------------------|--------------------------------|------------------------------------|--|--|
| Tracking ^d | 63.79 | 44.56 ^{***} (2.133) | 43.53 ^{***} (2.196) | -1.26 ^{***} (0.133) |
| Private school ^e | 17.56 | -0.11 (4.354) | -1.23 (4.002) | -0.47 (0.355) |
| Selective admissions | 30.17 | 12.31 ^{***} (2.344) | 12.78 ^{***} (2.312) | -0.81 ^{***} (0.242) |
| Selective residence | 35.39 | -6.70 ^{**} (2.824) | -7.81 ^{**} (3.082) | 0.19 (0.238) |
| Constant | | 684.84 ^{***} (203.405) | -56.31 (220.866) | 43.64 ^{**} (18.878) |
| Observations | | 17,252 | 17,252 | 17,252 |
| R-squared | | 0.736 | 0.674 | 0.158 |

Notes: Robust standard errors clustered at the country-level in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

^aAll regressions at school-level. Covariates in each regression include country dummies. The following controls are included: school-level average student age, gender, grade, whether the language of test is main language spoken at home, whether student lives with two parents, immigrant status, parents' immigrant status, index of socio-economic cultural status, school enrollment, dummies for community size where the school is located. School sampling weights are used. The analysis includes all the countries for which information on tracking is available. France excluded as the school questionnaire was not administered. 134 schools have no school-level variance as they have only one student.

^bA full description of variables is provided in the Appendix.

^cSchool fixed effects have been extracted from the student-level regression of test scores on school dummies and child/parents control characteristics.

^dThe tracking index is used as explanatory variable. The share of practice refers to the share of schools where age at first selection is below 16.

^ePrivate schools refer to the management of schools. Results related to the governance of funding are available from the authors.

scores among schools. The Appendix describes how we used the PISA questions to construct the variables that we use in ensuing regression analysis.

Table 3 examines the relation between these potential contributors to test scores and variation of scores within schools in two steps. The column labelled 'share of schools with practice' shows that almost two-thirds of the schools in the sample are in countries with tracking systems that move students to a higher level of school before the age of 16; that 18% of the schools are private; that 30% report selective admissions; and that 35% report giving heavy weight to enrolling students in the relevant geographic area.

The regression coefficients show the estimated impact of the various policies on three outcome measures: the average test score of a school (column 1), school fixed effects, which is the average score corrected for the characteristics of students (column 2) and the coefficient of variation in test scores within a school (column 3). The index of tracking (a country-level policy) has strong and consistent effects for all the

variables. It raises school test scores and the estimated school fixed effect and reduces the within school variation of scores – just what a tracking system should do. The private school variable has negligible insignificant effects and is not a major determinant of outcomes globally, though private schools may have substantial positive or negative relations to outcomes in some countries. The second most powerful statistical relationship is for the selective admissions policies, which like national tracking increases school test scores/estimated school fixed effects variables while reducing the coefficient of variation of within-school test scores.³⁰ Finally, the selective residence variable is associated with lower test scores/school fixed effects but has no effect on within-school variance. Measuring the share of variation in school fixed effects explained by the four sorting mechanisms as the difference in R^2 between the school fixed effects Equation (2), with and without the four sorting mechanisms, suggests that those factors account for a negligible amount of the variance in the fixed effects: dropping the variables reduces the R^2 by just 0.007 points. Still, the PISA questions are sufficiently weak and incomplete measures of sorting across schools, that the Table 3 analysis does not rule out sorting having larger contributions to the school fixed effects in the PISA test data.³¹ Moreover, for readers interested in sorting *per se*, our analysis leaves out a completely different channel of sorting: the assignment of students with differing abilities and interests to classes within schools (which is substantial in Canada and the United States – Oakes, 2005), that may affect the average level of scores and variance of scores within schools. For our purposes, the safest conclusion is that analysis of observable factors relating to the sorting of students among schools leaves a large variation in school fixed effects that could potentially reflect differences in school policies and practices, to which we turn next.

4.3. School policies and practices

This section seeks to identify specific school-level policies and practices *associated with* the substantial role of school fixed effects in the variance of student test scores found in Table 1. We choose the words ‘associated with’ rather than causes or determines because policies and practices are chosen by schools, for reasons about which the PISA provides no insight, rather than being randomly assigned by researchers or some natural experiment. This means that correlations could result not only from the effect of the policies/practices on test scores but also from the effect of test scores (or unmeasured attributes of students or schools associated with those scores) on selection

³⁰ Similar results are found in an alternative estimation when we include the PISA variable related to private funding.

³¹ We note that the selective residence question does not capture parental/student residential choices linked to school admission; none of the measures captures sorting of teachers; the private school dummy variable is a crude measure that does not distinguish between elite private schools and private schools for troubled students, nor for difference among schools in the public sector, as laid out by Nechyba (2011) and discussed briefly in Section 2.

of policies. Finding that certain policies/practices are highly related to test scores while others are not suggests directions for further work and areas for policy consideration. To the extent that schools select policies that work for their students, positive associations between policies and scores should reflect positive linkages in both ways and thus at the minimum identify the likely effect of the policy on scores.

The 2009 PISA data on school policies and teaching practices offer what is arguably the best cross-country school-based data to identify policies that work and those that do not. PISA obtains information on policies and practices from two sources. First, from principals who report on the presence of a large number of school characteristics and policies.³² Second, from students who report on the way they were taught in school. If either or both of these sets of information are related to the estimated school effects that dominate the PISA test scores we would have suggestive evidence that policies and practices matter and (conditional on likely dual causality) the ones that appear most promising for improving school performance.

Table 4 presents our analysis of the effect of school level policies or practices on the school fixed effects test scores for the 17,252 schools in the PISA data set. The column labelled 'Measures of School Policies or Practices' gives the specific policies and practices on which we focus. For ease of analysis, we grouped the school policies into three areas: *school selectivity*, *school autonomy and accountability*; and *school resources*, in which we include the attributes of teachers; and grouped the measures of practices into *approaches to teaching*, reported by students; and *staff attitude*, reported by principals. The Appendix provides the details on the questionnaire items we used.

For starters, the PISA data shows substantial differences in school policies and practices across countries as well as within countries (see Appendix Table A2). Principals in schools in Asian countries report having greater autonomy over curriculum and assessment but lower autonomy over resource allocation compared to principals in the OECD countries. Asian principals also report having greater instructional material, laboratory equipment and availability of computers and software for instruction and higher extra-curricular activities. Nearly 72% of Asian principals report academic selectivity in admissions compared to 33% of principals in the OECD countries.³³ And nearly all the Asian principals report that their teachers are fully certified with college degrees compared to 84% of the principals in the advanced OECD countries. Another major difference between these countries are that 44% of principals in OECD countries report that they are evaluated by achievement data and 43% report

³² We acknowledge the fact that there may be some measurement error or limited variability in some policy variables (e.g., Dustmann *et al.*, 2012). We also acknowledge the fact even though we tried to focus on those policies that are regarded as important determinants of educational attainment we are aware of the fact that we are not able to account for all the existing policies. We also looked closely at other sources (e.g., World Bank-Saber database, Eurydice reports) but given the nature of our database that includes both developing countries and advanced economies we could not find further consistent and comparable policy information.

³³ By contrast, 8% of Asian schools report ability grouping between classes within schools compared to 11% of schools in OECD countries.

Table 4. Regression coefficients of test scores and coefficient of variation of scores within schools on school policies and practices, PISA 2009^a

| | Test score all countries [<i>n</i> = 67] | Coefficient of variation of scores within schools | Test score countries by tracking | |
|--|---|--|--|--------------------------------|
| | | | No | Yes |
| | | | [<i>n</i> = 21] | [<i>n</i> = 46] |
| | (1) | (2) | (3) | (4) |
| Measures of school policies or practices ^b | | | | |
| Policies | | | | |
| Selectivity | | | | |
| Academic selectivity – [index based on factors for admittance; higher = greater selectivity] | 4.49 ^{***} (0.688) | -0.30 ^{**} (0.148) | 2.81 ^{***} (1.192) | 4.48 ^{***} (0.914) |
| Ability grouping between classes; [index derived from factor related to ability grouping] | -2.84 [*] (1.622) | 0.21 (0.150) | -2.98 [*] (1.632) | -3.18 (1.974) |
| Responsibility over curriculum and student assessment [index, mean = 0, sd = 1 for OECD] | -2.41 (1.673) | -0.07 (0.075) | 1.32 (2.378) | -3.54 ^{**} (1.747) |
| Responsibility over school resource allocation; [index standardized: mean = 0, sd = 1 for OECD] | -2.12 (1.337) | 0.15 (0.123) | -1.83 (2.795) | -2.34 (1.519) |
| Test info posted publicly [= 1 if posted publicly] | 4.06 ^{***} (1.902) | 0.01 (0.318) | 7.47 ^{***} (1.773) | 2.05 (2.406) |
| Achievement data used to monitor teachers' performance [dummy variable = 1 if use] | -3.73 ^{**} (1.856) | 0.46 (0.387) | -0.23 (1.731) | -3.22 (2.545) |
| Teacher peer review [dummy variable = 1 if teacher peer review for teaching/assessment] | -0.75 (2.362) | 0.01 (0.250) | -1.45 (3.133) | 0.24 (3.455) |
| External observer [dummy variable = 1 if external observers monitor teachers] | -2.56 (2.607) | 0.10 (0.205) | -8.11 [*] (4.455) | -0.47 (2.005) |
| Quality of educational Resources – [index based on items that can affect instruction at school] | 2.19 [*] (1.254) | -0.23 ^{**} (0.093) | 2.16 ^{***} (0.769) | 2.21 (1.722) |
| Extra-curricular activities [index based on the number of extra- curricular activities offered] | 1.28 (1.032) | -0.24 (0.150) | 2.86 ^{***} (0.948) | 0.28 (1.126) |
| Student-teacher ratio = enrolment divided by the total number of teachers | -0.26 ^{***} (0.102) | 0.02 ^{**} (0.008) | -0.36 ^{***} (0.133) | -0.22 ^{**} (0.121) |

Table 4. (Continued)

| | Test score all countries [<i>n</i> = 67] | Coefficient of variation of scores within schools | Test score countries by tracking | |
|---|--|--|--|------------------------------------|
| | | | No | Yes |
| | (1) | (2) | [<i>n</i> = 21] | [<i>n</i> = 46] |
| Measures of school policies or practices ^b | | | (3) | (4) |
| | | | | |
| Approaches to learning | % of <i>certified</i> teachers [the number of certified teachers by total number teachers] | −0.36 (3.146) | −10.70 ^{***} (3.360) | 2.89 (3.102) |
| | % of <i>qualified</i> teachers [dividing teachers with ISCED 5A by total number teachers] | 5.43 (6.076) | −8.80 (9.938) | 8.88 (6.501) |
| | Memorization [index on the use of memorization as a learning strategy] | −32.42 ^{***} (5.095) | −26.82 ^{**} (10.530) | −33.59 ^{***} (4.938) |
| | Control of learning strategies[index on use of comprehension checks as a learning strategy] | 56.00 ^{***} (5.330) | 49.77 ^{***} (8.355) | 59.34 ^{***} (6.788) |
| | Teachers' low expectations [dummy variable = 1 if principal said detrimental to learning] | 2.58 (2.761) | 11.64 ^{***} (2.896) | −1.49 (2.457) |
| Staff attitude | Constant | 522.76 ^{***} (150.039) | 328.23 [*] (188.307) | 625.79 ^{***} (173.761) |
| | Observations | 17,252 | 6,247 | 11,005 |
| | <i>R</i> -squared | 0.765 | 0.869 | 0.718 |

Notes: Robust standard errors clustered at the country-level in parentheses. Significance level: ****p* < 0.01, ***p* < 0.05, **p* < 0.1.

^aAll regressions at school-level. Covariates in each regression include country dummies. The following controls are included: school-level average student age, gender, grade, whether the language of test is main language spoken at home, whether student lives with two parents, immigrant status, parents' immigrant status, index of socio-economic cultural status, school enrollment, dummies for community size where the school is located. School sampling weights are used. The empirical analysis includes all the countries for which information on tracking is available. France excluded as the school questionnaire was not administered. 134 schools have no school-level variance as they have only one student.

^bA full description of variables is presented in the Appendix. Descriptive statistics are in Table A2.

using achievement data to evaluate school teachers. These numbers exceed the 18% of principals in Asian countries who say achievement data is used in their evaluations and the 34% who say they use it to evaluate teachers.

On the student side, questions regarding the extent to which students relate new information to prior knowledge and knowledge acquired in other situations, which PISA summarizes in an *index of elaboration*, is significantly lower in Asian countries. In addition, to our surprise Asian students report less use of memorization than their OECD counterparts. The *index of control of learning strategies* consists of techniques used by students to assess the effectiveness of their learning. It includes activities such as an understanding of what students need to learn, as well as self-supervision of learning and comprehension checks. It is on average higher among the OECD countries.

To see if any of these policies or practices affects test scores, we estimated Equation (1) with inclusion of variables for the policies and practices in place of school dummy variables, and with country dummy variables to pick up the large cross-country variation in the data.

Column 1 in Table 4 gives the regression coefficients linking the average school test scores to the policy and practice variables. Column 2 gives the coefficients for a similar equation linking the *variation in test scores within schools* on the policy and practice variables. Analysing the two columns together provides a way to probe potential causal links. For instance, factors associated with higher test scores and lower within-skill variability are likely to reflect sorting as well as other factors; while factors that affect levels of test scores but not the coefficient of variation are more likely to reflect the impact of practices or the effect of test scores on the choice of policy.

The results on academic selectivity policies of schools reproduce the pattern in Table 3: higher test scores and smaller within-school variance of scores for selective schools. The measure of within-school grouping of students by ability between classes also has opposite effects on the average test score and its variance but it reduces the average test score and increases within-school variance. Such a pattern could result from such groupings harming the lower-ability students by reducing their opportunity to learn from peers. But the pattern could also reflect the endogenous decision of schools to sort students by ability. Schools with weaker student bodies might be more likely to separate the more able from the rest. The PISA data do not allow us to differentiate these interpretations.

The coefficients for the school autonomy and accountability variables show some associations between policies and the level of scores (column 1) but essentially no relation between policies and the within-school coefficient of variation (column 2). The key positive relation for the level of test scores is in whether or not a school publicly posts information on student achievement data.³⁴ The biggest negative relation is in

³⁴ The achievement data includes 'aggregated school or grade-level test scores or grades, or graduation rates' (PISA 2009, School Questionnaire, p. 20).

using achievement data to evaluate principals, which may reflect school authorities relying more on achievement when schools are performing poorly than any adverse effect of this policy on educational outcomes.

The coefficients on the measures of school resources show that variables that reflect greater resources are positively associated with levels of scores and weakly negatively associated with the dispersion of scores. The most striking association is for the student to teacher ratio, which has a significant negative coefficient on test scores and significant positive coefficient on the variance of scores. Noting that the PISA measure is students/teachers rather than the more common teacher/student ratio, these results support the STAR finding that a higher teacher student ratio raises average performance rather than the negligible relationship found in the bulk of the literature cited earlier. A one SD change in the student-teacher ratio is associated with a 0.03 SD change in the test score. That higher teacher-student ratios also reduce the variance of scores makes it a 'win-win' in this context.

The most striking results in the table relate to the association between student reports on 'Approaches to Learning' and test scores. These are the strongest and most consistent patterns in the table. Memorization and use of elaboration are associated with low test scores (a one SD increase in memorization is associated with a 0.17 SD reduction in test scores; a one SD increase in elaboration is associated with a 0.09 SD reduction in test scores); whereas use of comprehension checks are associated with higher test scores (a one SD increase is associated with 0.27 SD increase in test scores). There are two possible reasons for such strong relations. One possibility is that teaching practices that stress comprehension rather than memorization produce better results on the PISA tests that stress use of knowledge in practical settings. The other possibility is that the correlation reflects schools with better students eschewing memorization and elaboration techniques in favour of comprehension checks. With respect to the within-school variance of test scores, the estimated coefficients show that: memorization and elaboration are associated with greater variation while comprehension checks are associated with lower variation. Finally, there are no clear patterns in the relations between the principal reports of staff attitudes and test scores or variance in scores within schools.

To get a better sense of the robustness of the estimated patterns across countries with different tracking policies, we divided the sample into groups: schools in countries with a tracking system (defined by the dichotomous tracking variable described in Section 4.1 and those without such a system. Since schools without tracking deal with more heterogeneity in their student bodies, they are likely to rely on the policies and practices listed in the table than those with tracking, possibly with different effects. In fact, test scores in the schools without tracking appear to be more responsive to policies than test scores in schools with tracking and sometimes with differently estimated signs on the variables. For instance, the posting test score variable has a large positive effect for schools without tracking compared to a smaller insignificant scores for schools with tracking while four of the measures of school resources obtain statistically significant coefficients for the non-trackers compared to a single measure – the

student-teacher ratio – for schools with tracking. The overall pattern shows up in the noticeably higher R^2 in the column 3 regression than in the column 4 regression. Whether the causal pattern runs from the test scores to the policies or from the policies to the test scores, the differences between the two groups are a warning sign that similar policies may be associated with different outcomes depending on other characteristics of the school system and country.

One set of variables in Table 4, however, shows no relation between schools in countries with tracking and schools in countries without tracking. This is the ‘approaches to learning’ variables that measure the way students say they are being taught. Here, the variables have similar highly significant effects regardless of national tracking policies. This robust result suggests that the mode of teaching may be the key to understanding the strong school effects in the data. Future research into school practices may profit from investing more in obtaining student views of how the school operates than in giving extended questionnaires to principals.

A parallel way to examine the extent to which school fixed effects are associated with what schools do as opposed to the sorting of students among schools is to examine directly the link between estimated fixed effects and policies and practices. To do this we run the following regression:

$$\beta'_{3sc} = \alpha_1 \mathbf{S}_{sc} + \varepsilon_{sc} \quad (3)$$

where β'_{3sc} is the vector of estimated fixed school effects for school s in country c and \mathbf{S}_{sc} is the vector of school policies and practices in school s in country c .

Table 5 summarizes the results from estimating variants of Equation (3). Since the estimated coefficients on specific policies and practices are similar to those in Table 4,³⁵ the table gives the residual variances from the regressions and the summary R^2 statistics for different specifications. These estimates tell us the extent to which the whole set of policies and practices help explain the fixed effects. The line labelled school fixed effects gives the total variance in the estimated fixed coefficients. Line 1 shows that with no other variables in the equation, the vector of policies and practices by itself accounts for 44.3% of the variance. Line 2 shows that the country dummies account for 52.6% of the variance in the fixed effects while line 3 shows that addition of the vector of policies and practices raises the R^2 to 63.9%. That is, the policies and practices explain 23.8% $[=(0.639 - 0.526)/(1 - 0.526)]$ of the variance beyond that associated with country differences.

The remaining lines in the table explore the effects of adding school-level covariates to these calculations, where the covariates are average levels of student attributes as defined in the table note. Line 4 shows that these attributes by themselves account for 21.4% of the variance. Addition of policy and practices in line 5 raises the R^2 to 0.583, so that policies and practices explain 46.9% $[=(0.583 - 0.214)/(1 - 0.214)]$ of

³⁵ The results from these regressions are available from the authors on request.

Table 5. School fixed effects and residual variances for different specifications^a

| | Variables | Residual variance in estimated school fixed effect(1) | R^2 (2) |
|-----|--|--|--------------|
| | School fixed effects ^b | 5684.050 | |
| [1] | Policies and practices ^c | 3166.016 | 0.443 |
| [2] | Country dummies ^d | 2694.240 | 0.526 |
| [3] | Policies and practices with country dummies ^e | 2051.942 | 0.639 |
| [4] | School-level covariates ^f | 4467.663 | 0.214 |
| [5] | School-level covariates with policies and practices ^g | 2370.249 | 0.583 |
| [6] | School-level covariates with country dummies ^h | 2392.985 | 0.579 |
| [7] | School-level covariates with policies and practices and country dummies ⁱ | 1659.743 | 0.708 |

^aResults presented in the table are derived from school-level regressions (total number of schools = 17,252). The dependent variable is school fixed effects. France excluded as the school questionnaire was not administered. School sampling weights are used. Analysis based on clustering-robust standard errors.

^bSchool fixed effects have been extracted from a student-level regression which includes student background characteristics and school dummies.

^cThe regression includes as explanatory variables school policies and practices. These are the policies and practices presented in Table 4 and described in the Appendix.

^dThe regression includes as explanatory variables country dummies.

^eThe regression includes as explanatory variables school policies and practices as described above, and country dummies.

^fThe regression includes as explanatory variables school-level covariates. Covariates include the following variables: school-level average student age, gender, grade, whether the language of test is main language spoken at home, whether student lives with two parents, immigrant status, parents' immigrant status, index of socioeconomic cultural status, school enrolment, dummies for community size where the school is located.

^gThe regression includes as explanatory variables school-level covariates as described above, and school policies and practices.

^hThe regression includes as explanatory variables school-level covariates as described above and country dummies.

ⁱThe regression includes as explanatory variables school-level covariates as described above, school policies and practices, and country dummies.

the variance beyond that associated with other school level factors. Finally, lines 6 and 7 shows that addition of school policies and practices to regressions with covariates and country dummies raises the R^2 from 0.579 (line 5) to 0.708, so that policies and practices explain 30.6% $[(0.708 - 0.579)/(1 - 0.579)]$ of the variance beyond that associated with other school level factors and country dummies.

In sum, the calculations in Table 5 show that fixed school effects are sufficiently related to school practices and policies to reflect more than sorting of students. Lacking exogenous variation in policies and practices to separate the link from the policies and practices to test scores from the endogeneity of policies and practices to student attributes associated with test scores, we cannot determine the magnitudes of the policy factors on school level test scores. What the association between the fixed effects and policies and practices does is to direct attention to school effects and decisions as opposed to sorting of students in explaining the variance of test scores and thus in developing school-level policies to improve scores.

5. SCHOOL EFFECTS AND FAMILY BACKGROUND

As noted in the Section 2 literature review, studies invariably find that family background factors are strongly associated with student performance in schools and test scores. Table 1 showed that background factors by themselves explained 33% of the variance in test scores among all students and explained from 14% (Asian countries) to 29% (EU15) of the variance in test scores within country groups. Given the large role of the school which students attend in the variance of test scores, we ask the extent to which, if at all, the schools which students attend mediate the relation between family background and student test scores. The answer is a surprisingly large extent.

The PISA contains three potential measures of family background:³⁶ the education of their parent,³⁷ the number of books in the household,³⁸ and a composite 'index of socio-economic and cultural status' (ESCS)³⁹ developed by PISA. Students report on their parents' education and on the number of books in their household. PISA calculates the ESCS based on the level of parental education converted into years of schooling,⁴⁰ an index of parental occupation, an index of family wealth, cultural possessions, home educational resources, and an index of home possessions, which includes the variable number of books at home.

Table 6 records the mean and standard deviation (in parentheses) of these variables among students in all PISA countries and in the five groupings that we have examined. The means for the different groupings show variation in the levels of the background variables while the standard deviations show variation within the country groupings.

Students from Asian countries have a higher share of parents who have completed higher education (39%) compared to the students in the OECD countries (31%). Parental education is almost always positively associated with better school performance by children.⁴¹ Asian students taking the PISA tests also report living in families with more books than students in OECD countries: 43% of Asian students live in

³⁶ See Schuetz *et al.* (2008) for a discussion of the advantages/disadvantages of these different measures.

³⁷ Answers are based on the ISCED classification (PISA 2009, Student Questionnaire, questions n. 10, 11, 14, 15). PISA derives an index of the highest level of education of either parent. It also converts this into the highest number of years of schooling.

³⁸ This variable comes from students' responses to a question about the number of books in their home exclusive of magazines, newspapers and schoolbooks. Students can choose one of the six categories: 0–10, 11–25, 26–100, 101–200, 201–500, more than 500 (PISA 2009, Student Questionnaire, question n. 22).

³⁹ The OECD Technical Report (OECD, 2012a.) details the construction of this index. The index is standardized to have an OECD mean of 0 and a standard deviation equal to 1.

⁴⁰ The variable is derived from the student questionnaire, in which PISA asks students the highest level of schooling completed by their father/mother and whether their father/mother has completed undergraduate or postgraduate degrees.

⁴¹ Some analysts argue that parents' education increases the marginal productivity of children's education, by raising expectations about educational success (Davis-Kean, 2005; Bjorklund and Salvanes, 2010) but it can also operate through the higher income associated with higher parental education.

Table 6. Mean of family background measures, PISA 2009

| Country group ^g | Years of schooling ^a | Books at home ^c | Index of Economic, Social and Cultural Status (ESCS) ^e |
|----------------------------|---------------------------------|----------------------------|---|
| | Mean ^b | Mean category ^d | Mean ^f |
| All | 12.17 (3.63) | 2.82 (1.42) | -0.54 (1.21) |
| OECD | 13.11 (3.11) | 3.16 (1.45) | -0.03 (1.01) |
| EU 15+ ^h | 13.45 (2.98) | 3.32 (1.44) | 0.02 (0.93) |
| Asian | 13.56 (2.39) | 3.45 (1.37) | -0.15 (0.82) |
| Other high income | 13.28 (1.77) | 3.30 (1.36) | -0.20 (0.82) |
| Middle income | 10.77 (4.05) | 2.26 (1.18) | -1.18 (1.22) |

Note: Sampling probability weights have been used in order to provide descriptive statistics representative for each country.

^aHighest level of education of either parent according to the ISCED classification. Parental education has been converted into years of schooling according to a country-specific scale provided in PISA 2009 (Vol. I, Table A1.1, p. 168) [completed primary education (ISCED 1); completed secondary education (ISCED 2 (lower secondary), ISCED Level 3B or 3C (vocational/pre-vocational upper secondary); ISCED 3A (upper secondary) and/or ISCED 4 (non-tertiary post-secondary)); completed tertiary education, that is theoretically oriented tertiary and postgraduate (ISCED 5A, 6)].

^bAverage years of schooling with standard deviation in parentheses.

^cBooks at home shows the share of individuals according to the number of books they report having in their home.

^dMean of the books-at-home categories (1–6). These are the six categories: 1 = 1–10 books, 2 = 11–25 books, 3 = 26–100, 4 = 101–200, 5 = 201–500, 6 = more than 500 books. Standard deviation in parentheses.

^eThe ESCS index is constructed by using the following three sub-components: the higher parental occupation index, the higher parental education index and the index of home possessions.

^fMean of the ECSC index. Standard deviation in parentheses.

^gOnly a low-income country participated in the PISA 2009 study: Kyrgyzstan, so we did not include this group. The analysis has been carried out for this country and is available from the authors upon request.

^hThis category includes Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom, Norway and Switzerland.

families with more than 100 books compared to 34% of students in the OECD country sample. The ESCS index gives students in the EU15+ the highest score and students in middle income countries the lowest scores.

We have examined the mediating role of school fixed effects in the relation between each of the measures and test scores by estimating variants of Equation (1).⁴² We first estimated the effect of a given background factor with no measure

⁴² This production function approach is commonly used in the literature (Hanushek and Woessmann, 2011a). The main underlying assumptions are related to the fact that the production function is known and inputs are used to maximize student learning outcomes. Hanushek (1979, 1986) describes some of its limitations. Schuetz *et al.* (2008) note that country-specific results may be affected by the measurement of background factors.

of school effects in the equation and then estimated the equation with the inclusion of fixed school effects.⁴³ We obtained similar results for each of the three background measures.⁴⁴ For simplicity we summarize our findings for the ESCS variable, which has the virtue of encompassing parental schooling and books at home measures of background.

Table 7 gives results for all countries and for countries with early tracking policies and countries without them, the better to isolate effects within school unrelated to tracking. Results for particular country groupings are available from the authors. The table records coefficients on the ESCS measure of background and the R^2 associated with the regression of test scores on the variables specified in the row and in the table note. Row (1) records the estimated coefficients on family background with country fixed effects as the only covariate so that the analysis focuses on within-country variation in test scores. Row (2) includes a set of student-level and a set of school-level controls that reflect in addition differences in the urban location and size of schools across countries (see Appendix Table A1). In both rows the coefficients of family background are large and significant. In row (1) for the all country sample, a one-SD change in the ESCS index is associated with a 0.36 SD change in test scores. The addition of background and school control variables reduces the estimated size of the ESCS effect to 0.28 SD. But it is in row (3), where we add the school dummy variables that the coefficients on the ESCS background measures drop greatly. For all countries, the coefficient drops from 24.20 to 10.28 – a 58% fall. The effect size falls to 0.16 SDs. While the extent of the decline differs among the groups, the calculations in ensuing columns show that it is large in all cases.⁴⁵

PISA contains an alternative set of measures designed to cast light on the relation between family background and test scores. Fourteen countries⁴⁶ in the 2009 wave of PISA practices administered a parent questionnaire that asked about parenting practices relevant to education.⁴⁷ The OECD analysis of these questions (OECD, 2012b) shows that parental activities are positively correlated with test scores but that specific activities with the child have different relations to outcomes. A parent discussing what their child is doing is positively related to test scores whereas a parent discussing their

⁴³ Dustmann *et al.* (2012) used a similar approach to examine the effect of school resources, class size and accountability on the gap in learning outcomes among migrant and native students in PISA 2006.

⁴⁴ These results are available from the authors upon request.

⁴⁵ Table 7 only shows the coefficients related to the family background variables. The complete regression results are available from the authors.

⁴⁶ These are the countries: Hong Kong, Korea, Macao-China, Chile, Denmark, Germany, Hungary, Italy, New Zealand, Poland and Portugal. The parent data file includes the variables derived from the parent context questionnaire (OECD, 2012a, p. 328). The Parent Questionnaire of the PISA 2009 study was delivered to parents by students who had to return it to school once it had been filled in. At page 2 of the Parent Questionnaire it is explained that 'This survey should be completed by a parent (or jointly by both parents) or other <primary caregiver> of the student'.

⁴⁷ PISA 2006 also included some parent questions. Before the 2006 and 2009 waves, cross-country data sets of student performance did not provide much information on parents' participation in home and school-related activities.

Table 7. Regression coefficients of Student test scores on ESCS measure of family background PISA 2009^a

| Country group [number of observations ^b , ^c] | | All | | Without school tracking ^d | | With school tracking ^d | |
|--|--|---------------------|----------------|---|----------------|--------------------------------------|----------------|
| Empirical model | | ESCS coefficient | R ² | ESCS coefficient | R ² | ESCS coefficient | R ² |
| [1] | Test score on ESCS ^c (with <i>country fixed effects</i>) | 31.21*** (0.536) | 0.42 | 31.48*** (1.030) | 0.41 | 31.09*** (0.609) | 0.42 |
| [2] | Test score on ESCS ^c with other covariates ^f , <i>country fixed effects</i>) | 24.20*** (0.497) | 0.48 | 25.62*** (0.899) | 0.48 | 23.05*** (0.572) | 0.49 |
| [3] | Test score on ESCS ^c with school fixed effects and other covariates ^f | 10.28*** (0.317) | 0.66 | 13.02*** (0.644) | 0.62 | 8.53*** (0.309) | 0.69 |
| [4] | Test score on ESCS ^c with parental involvement indicators ^g , other covariates ^f , <i>school fixed effects</i>) | 9.39*** (0.51) | 0.64 | 15.80*** (1.07) | 0.44 | 8.88*** (0.55) | 0.65 |

Notes: Regressions weighted by the students' sampling probabilities.

^aEstimates using parents' education and number of books at home available from the authors upon request.

^bNumber of observations in [4] for the 14 countries that administered the parental involvement questionnaire. Regressions [1]–[3] estimated for the 14 country sample given similar results to those in the table. Results are available from the authors. Detailed descriptive statistics related to the parental involvement variables are available from the authors.

^cThe number of observations for the different groups of countries and related estimation is the following: All: ([1]–[3]: 515,956; [4]:101,370); Without school tracking ([1]–[3]: 171,885; [4]: 21,356); With school tracking ([1]–[3]: 316,357; [4]: 80,014).

^dThe 'without school tracking' group includes countries where age at first selection of students into different tracks is equal to 16 or higher. The group 'with school tracking' includes countries where age at first selection is below 16. Information on school tracking: sources used to derive this variable are described in the footnote of Table 2. The country-specific information on school tracking is available from the authors.

^eThe ESCS index is constructed from the following three sub-components: the higher parental occupation index, the higher parental education index and the index of home possessions.

^fCovariates: student age, gender, grade, whether the language of test is the main language spoken at home, whether the student lives with two parents, immigrant status, parents' immigrant status, school enrolment, dummies for community size where the school is located.

^gParental involvement is measured with four indices extracted from the parental questionnaire described in the appendix: 'parental involvement', 'resources', 'early support at ISCED1', 'current support'.

Significance level (based on clustering-robust standard errors): * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

child's progress in school with a teacher is negatively related to test scores, presumably because the parent is responding to the student not doing that well in school. The OECD also finds that parental involvement is higher in more affluent households.

However, it warns that the high non-response rate in some countries raises questions about the results (OECD, 2012a, p. 328).⁴⁸

We selected from the parent questionnaire four areas of involvement to examine: early parental investment in children exemplified by questions about how often they read books with the child at an early age; participation in school-related activities such as whether parents participate in the local school governance; provision of resources to the child such as subscription to a journal or magazine; and current activities with the child such as discussing political or social issues with the child.⁴⁹ Row (4) in Table 7 shows the addition of the parental involvement indexes had a modest effect on the estimated impact of the ESCS background measure overall, decreasing it among OECD countries to an estimated effect size of 0.09 of a standard deviation, while increasing the background coefficient among Asian countries.⁵⁰ It does not have the consistent effect on the estimated background effect that the schooling fixed effects have.⁵¹

⁴⁸ Some bias may also be generated by the design of the survey: better educated parents may be more likely to fill this optional questionnaire and more diligent students – with higher test scores – may be more likely to bring the questionnaire home, remember to ask their parents to fill it and after to return it to school (OECD, 2012b, p. 91). In this regard, descriptive statistics presented in the OECD study (2012b, p. 163) carried out for the entire PISA 2009 sample and for the sample of students who answered the parent questionnaire show a significant heterogeneity in terms of response rate across countries. Apart from Poland where the questions on parental involvement were not asked (OECD, 2012b, p. 50), the response rate varies from about 100% in Macao-China to about 60% in Denmark. There are also some differences in terms of demographic characteristics. Students in the sample with answers to the parent questionnaire have on average better reading performance, a higher socio-economic status, are more likely to be a girl and less likely to be of immigrant status. Moreover, a difference between the information on parents' education as reported by students and parents themselves should be acknowledged, being information self-reported by parents on average than the one reported by their children (Jerrim and Micklewright, 2012).

⁴⁹ We use indexes based on the parents' answers to the questions in each area available in the PISA database: *Early Parental Involvement Index* of activities with the child (when the child attended the first year of ISCED 1); *Current Parental Involvement Index* of current activities with the child; *Parent Participation in School Related Activities Index* of participation in school-related activities; *Resources Directly Available to the Child Index* of resources available to the child. Section A2 in the appendix provides a detailed description of the indexes, items in the questionnaires on which they are based and how they have been constructed. The PISA 2009 Technical Report explains the computation, scaling methods and weights involved in the construction of the indexes (OECD, 2010d, 2012a). The PISA variables represent a limited selection of the many aspects and activities that shape the relationship between parents and children and could impact on the students' learning progress. Descriptive statistics related to the parental involvement variables are available from the authors upon request.

⁵⁰ Rows 1–3 include all countries in the specified group. Row 4 includes only the 14 countries where the parental questionnaire was administered. Regressions in rows 1–3 have also been estimated with the same sample as in row 4. Results show that parental involvement has only a limited impact in reducing the ESCS coefficient for all country groups. These additional results together with the results for specific groups of countries are available from the authors upon request.

⁵¹ Our findings suggest that measures of parental practices do not account for much of the parental background factors. Instead the major factor that reduces the impact of background factors on outcomes are school fixed effects. It is important to notice that given the structure of the PISA database we are able to control for a limited number of family background characteristics and parental practices and we are not able to control for the unobserved parental characteristics. Therefore, given the existing asymmetry of our analysis we are not able to draw conclusions on the role that unobserved parental characteristics correlated with school fixed effects may play in affecting outcomes.

Measured school policies and practices have a greater effect in mediating the influence of background on test scores than measured parental involvement activities.

6. CONCLUSION AND POLICY IMPLICATIONS

Our analysis of the 2009 wave of the PISA data set has found that much of the variation of test scores is associated with the school students attend and that school fixed effects are a major pathway for the link between measures of family background and test scores.

Within the limits of observational data, we estimated the extent to which the variation in student performance across schools reflects mechanisms that sort students into different types of schools as opposed to the relation between school policies/practices and student performance. We examined national policies toward tracking students at different ages, school policies toward selective admissions and toward admissions based on residency and between private and public schools and found that while greater sorting by country or school raised the variance of test scores among schools and lowered variance of test scores within schools, measurable sorting leaves a sizeable proportion of school fixed effects for school educational policies and practices to affect student performance and thus for policies and practices to improve student test scores.

Using the reports of principals and students on the PISA, we examined the association between explicit policies/practices and the test scores of students attending different schools.

Consistent with our analysis of sorting, the full set of measured policies and practices accounted for a sufficient proportion of the school fixed effects to support the view that policies and practices can affect test scores in important ways. We found that several policies, particularly those relating to school resources, were significantly associated with test scores, though some had differing effects on schools with different tracking policies, highlighting the potential for the same policy to have different effects in different settings. Since principals undoubtedly differ in how they implement policies across countries and within countries, we believe that new PISA questions on implementation would increase the proportion of the test score variance associated with policies and practices.

Our analysis of the role of sorting as opposed to within-school policies and practices has implications on the continual debate regarding these alternative ways of improving educational outcomes. As we saw in the literature review, much research using cross-country data focuses on the effects of explicit national tracking on outcomes while recognizing that other mechanism also sort students with particular abilities into schools with students like themselves. Our data shows that, taking sorting as given, there is still a large space for school level policies and practices within sorting regimes. Schools in countries which track students early and schools that admit students with greater selectivity show substantial school effects independent of the sorting and measured student attributes. Within sorting groups, moreover, schools operate with

different policies and practices, some of which are associated with different test score outcomes that suggests substantial potential for improvements in scores within sorting groups.

Our Table 4 analysis found that while the approaches to learning practices are associated with students' test scores in a similar way between groups of countries with tracking and those without tracking, as defined in the tracking literature, other policies such as those relating to school autonomy and accountability are associated with test scores differently in tracking than in non-tracking environments. To the extent that these effects reflect the differential impact of policies on students, they direct policy attention at the value of developing different school policies depending on the extent of national tracking. Reducing school autonomy in countries without sorting is associated with poorer performance while it has modest negligible effects in countries with sorting. The key policy implication is that the choice of some within school policies and practices should be considered differently depending on sorting and that international evidence that something works in a system with tracking cannot be taken as indicative that it will work in a system without tracking.

The biggest problem in analysing the PISA evidence is unpacking the causality of the observed associations: the extent to which they reflect the effect of policies on test scores or of student performance and attributes on policies. We suggest that the best way to untangle this is to address it directly on the principals' questionnaire: to add a set of questions asking when and why a policy was adopted and whether the school had assessed it through some sort of before/after or other analysis. If school policies matter and effectiveness depends on implementation, the best way for international tests to move from measuring outcomes to helping devise reforms in schooling would be to measure these factors at the school level.

The strongest and most consistent relation between policies and practices in the PISA data was not however, based on principal reports but on student reports on teaching practices at their school. The strength and consistency of the relation between student reports and school test performances suggest that the most promising way for PISA and other international surveys to increase our understanding of school effects would be through obtaining more information about school experiences from students themselves. A set of questions to students on school practices and individual experiences would offer a new line of sight into what goes on in schools that actually affects student performance.

Following this line of thought, the PISA (or other surveys) could illuminate sorting mechanisms by asking students or parents about why the student was attending the current school and how they were admitted. Bratti and Checchi's (2013) finding in an analysis of a Italian re-test of schools in the 2009 PISA that cross-sectional test scores were good measures of changes in student performance when student attrition was low but bad measures when student attrition was high suggests the value of questions on how long students were in a given school.

While it can only be suggestive for directions of policy reforms, our analysis of the school fixed effects in the PISA data suggests that policies related to the quality of educational resources and practices at the classroom level are good candidates for improvements in performance. We suspect that the reason they turned out to be strongly linked to school test performances is that they are closer to actual classroom activities than some of the accountability and autonomy managerial practice questions and to the parental activity questions on which the PISA focused. Still, even finding avenues for improvement does not itself justify schools seeking those reforms. It is necessary as well to estimate the cost of the reforms (Hanushek and Woessmann, 2011b) and the length of time it takes for reforms to become effective (Braga *et al.*, 2013).

Still, given the great success of PISA in discussion of educational outcomes around the world and its influence on policies, we believe that improving it in ways that will better illuminate school effects and help identify the causal forces behind those effects has the potential for catalysing school level reforms that could improve the performance of poorer performing schools and reduce the magnitude of the schools fixed effect that we found in the 2009 PISA.

Discussion

Daniele Checchi

University of Milan

The paper by Freeman and Viarengo raises important questions that any educational authority would like to know the answer to, in order to manage important school decisions, like student allocation and/or teacher allocation to schools, in order to maximize school effectiveness at national level. However, my impression is that they aim to obtain more information from PISA than these data can provide. Given their extended presentation, I will only recall here that PISA consists of a representative cross-section of 15-year-old students in formal educational institutions.⁵²

Let us start with their first results. In all countries and groups of countries, they find that a substantial proportion of the variation of test scores within countries is associated with attending a specific school. This is a well-known fact in the sociological literature, which speaks of school stratification. From an economic point of view, it implies that schools can be heterogeneous, due to either

⁵² This may be problematic in countries where students can be retained, since repeated failure may retain students in lower secondary schools. In addition, early school leavers are also excluded from the test, thus artificially raising the scores in countries with the highest drop out rates.

- teaching practices (if students were randomly allocated to schools, every difference would be teaching practices), or
- sorting (if schools were to use the same teaching practices, every difference among them would be attributable to sorting, either of students and/or of teachers).

In principle, these two alternative situations could be identified iff one had an exogenous source of variations for one or the other. The authors intend to use the tracking age as such an instrument, but it is just one source of practice. They add selection in admission as a second one. But the list is longer (think of private/public).

From a conceptual point of view this distinction is hard to conceive, because sorting may occur *because of* different teaching practices.

Think of a situation with heterogeneous students of mass one; their ability is uniformly distributed in $[a, \bar{a}]$. There are just two schools, A and B . The former is talent oriented, the latter specializes in weaker students. Performance in school is given by $y_{ij} = \alpha_j + \beta_j a_{ij}$, with $\alpha_A > \alpha_B$ and $\beta_A < \beta_B$, where y_{ij} is the educational outcome of student i when attending school j . If teaching practices are common knowledge, the students will sort according to their level of ability: students with $a > \frac{\alpha_B - \alpha_A}{\beta_A - \beta_B}$ will attend school A , while the others will go in school B .

The average performance of school A will be

$$\bar{y}_A = \alpha_A + \beta_A \frac{\bar{a} + \hat{a}}{2}, \quad \hat{a} = \frac{\alpha_B - \alpha_A}{\beta_A - \beta_B}$$

whereas the average performance of school B is

$$y_B = \alpha_B + \beta_B \frac{a + \hat{a}}{2}$$

We could in principle separate the role of teaching practices if we could observe the ability of students. We actually do not necessarily need ability, we would like to observe the underlying expectations that drive students' sorting. But in my opinion nothing of the sort is available in PISA, since the rich array of variables describing parental background does not necessarily capture all individual heterogeneity.

The authors believe it is possible to identify the relative contribution of students sorting by using variations in the tracking policies of different countries. I have two objections to this strategy:

- tracking is not exogenous, but reflects the prevailing opinions of the local constituency with respect to school differentiation and/or desired educational achievements in the population. When some German *landers* retained that early tracking was detrimental to student achievements, they simply introduced comprehensive schools (*gesamtschule*).

- tracking is not only described by age tracking; one may also want to consider the number of tracks available, the share of students who are actually tracked, the degree of differentiation of educational contents, the level of qualification of teachers in different tracks, the absence/existence of the dual tertiary system, and so on. Even if tracking is a general institutional feature, the selection of alternative variables makes a difference (at least from an econometric point of view).

However, the authors are ingenious in exploring new measures of the degree of school stratification, by estimating the relationship between school fixed effects and (within-country) school rank. Conditional on an equivalent number of schools in each country, this relationship (Equation 2 in the paper) offers an indication of how strong is the sorting effect, which comes out correlated with the extent of tracking (measured by age of student first allocation to tracks). These correlations do not tell us the direction of causality. Moreover, they do not exhaust the potential sources of student sorting. Consider, for example, family income as another source of student sorting when private schools are available. Think of religion and/or ethnicity, when cleavages are endemic in some society. If one would repeat the regression presented in Table 2 with one or more variables in this list, he would probably get similar results. The authors go in this direction with the results of Table 3, where surprisingly they find no significance in the correlation between school fixed effect and the extent of the private sector in education provision within the country.⁵³

The third contribution of the paper is the investigation of teaching practices. This is an endless effort, because information collected from students is abundant but confused. If we look at Column 1 of Table 4 we find some results that are at odds with common wisdom. Most of the variables on school autonomy and accountability are insignificant (or wrongly signed, as in the case of principal evaluation – compare this finding with Woessmann *et al.*, 2009). Similar results are obtained for school resources, while emphasis is placed on student learning strategies. But this factor is extracted from answers to elementary questions (like ‘when I study, I try to memorize everything that is covered in the text’), on which students have to express their relative agreement or disagreement. In Table 4 the authors find a negative correlation between school fixed effect and *memorization* and *elaboration*, but positive with *control of learning strategies*. What is surprising is that when we consider the raw data at student level, all these variables are positively correlated.⁵⁴ My impression is that students are rather confused when describing their learning strategies, and one should explore the

⁵³ This may due to multicollinearity between being private and selecting students for admissions. In Bertola and Checchi (2013) we have shown that student sorting may go in either direction (best students sorted into the private, or best students sorted into the public sector), depending on teaching techniques. This may account for the non-significance of the coefficient on private found by the authors.

⁵⁴ Given 506,640 observations, the correlation between *memorization* and *elaboration* is 0.40, between *memorization* and *control of learning strategies* is 0.52, while the correlation between *elaboration* and *control of learning strategies* is 0.55.

degree of disagreement within the classes before making inference for educational outcomes at school level.

Marco Manacorda

Queen Mary University of London

This paper is deemed to have a large echo among policymakers and researchers in education alike. Based largely on observational data, the authors makes a quite bold – and ultimately I believe convincing – statement about the major role of school policies and practices in affecting children's school outcomes.

Ultimately, the paper provides some reason for optimism as it suggests that there is room for policymakers to improve learning by fostering the adoption of good practices and policies by schools, something that seems easier to achieve, at least in the short run, than affecting parental behaviour, leaving aside parental characteristics.

What makes the paper extraordinary – and the empirical exercise equally daunting – is the sheer amount of data covering 74 countries and providing very detailed information on school policies and practices as well as information (but only for a sample of countries) on parental involvement at home and in school.

I thought that the best way to discuss the paper was by trying to 'unpack it' with a policy audience in mind. Indeed the paper is so rich that it is at times slightly hard to digest even for a consumed data practitioner like myself.

The difficulty that anybody wanting to address the question of the role of school versus parental inputs in affecting school outcomes is that children are not randomly allocated across schools. Showing that children in 'better schools' do better in terms of test scores raises the suspicion that this is not due to what schools do but to the fact that children who would have done better anyways (from more affluent backgrounds, with more motivated parents, etc.) also attend schools that are better managed, with more and better quality inputs, better teachers etc. In other terms students *sort* across schools.

The basic idea of the paper to get round this problem, with a clever twist that I discuss below, is quite simple. Via means of regression analysis, the authors decompose the variation in test scores into a component that is due to observable parental inputs and another component that is due to observable plus unobservable school inputs. The latter are subsumed by school fixed effects. One can try to assess the role of observable school practices and policies in explaining differences in test scores across schools by examining how school fixed effects vary with such policies and practices.

One has to be cautious here: as long as there is sorting along unobservable parental and child characteristics, school fixed effects or observable practices will soak up some of the effect of such unobservables, and one will most likely end up attributing too much weight to school inputs.

That sorting indeed matters is clear from the results in Tables 2 and 3. The logic of these tables is stringent: if sorting partly happens through early tracking, and tracking simply reshuffles children across schools with no direct effect on learning, then the dispersion in average scores among schools will be larger in countries that track children earlier. By the opposite token, the dispersion in test scores within schools will be smaller. This is precisely what is shown in Tables 2 and 3. This suggests that sorting matters although tracking (as well as other selective policies) does not seem to explain much of the variation in test scores across schools. This is reassuring, although the concern remains that the measures of sorting available do not explain much of the actual sorting that in practice occurs (through residential segregation for example).

Table 4 is to me the most interesting table in the paper. Column 1 shows that school policies and practices matter in explaining differences in test scores across schools. As said, sorting remains potentially an issue as school policies and practices might be correlated with unobserved individual or parental characteristics. Where the table is really convincing, and the approach very smart while remaining simple, is when it contrasts column 1 and column 2. The latter examines how the dispersion in individual test scores within schools (as opposed to the dispersion in average test scores across schools in column 1) is affected by such policies and practices. The authors' argument is that if one finds that some policies affect both average test scores and the dispersion of test scores, as in the case of tracking discussed above, then the founded suspicion is that the effect of these policies and practices works through sorting. However, if policies and practices affect the dispersion in average scores across schools but not the within school dispersion in test scores, then this brings ammunition to the argument that these practices are truly affective in improving outcomes. Indeed, it appears that, in most cases, these observable policies and practices affect average outcomes (column 1) with little or no effect on within schools dispersion (column 2).

With this conclusion in mind, Table 5 finally attempts to put a number of how much measurable school policies and practices matter in explaining the variation in test scores across schools. The authors estimate this number at between 23% and 30%. This is perhaps slightly overoptimistic, given that sorting remains still an issue, although based on the evidence in Table 4, the bias is presumably small.

In sum, the authors have proved creative in trying to come to credible conclusions based on purely observational data. The paper makes a rather compelling case for some – but not all – school policies and practices to be a major contributor to differences in achievement across children.

One additional consideration that comes from reading this paper is that greater effort should be exerted to collect additional information in the PISA data. Although important steps have been taken in this direction, it appears that additional information about parental involvement (for all countries) and characteristics, children's effort and detailed residential location would prove extremely valuable to advance our

understanding of the determinants of children's school outcomes across and within countries.

Panel discussion

Fabrice Murtin found the central result of the paper regarding school effects to be credible. However, he contended that a major factor influencing the quality of a school is the type of principal leading it (e.g. level of experience, powers/authority). Murtin explained that this is a particularly relevant driver in disadvantaged areas in OECD countries and suggested further exploring this venue. Second, he noted that school autonomy is not expected to have an impact on its own. Murtin advised the authors to interact autonomy with school accountability in order to potentially extract a statistically significant effect. Ester Faia stated that the identification problem is prevalent in a number of literatures. She mentioned that Becker stipulates the conditions for identifying a theoretical model (unless one assumes the presence of some form of frictions).

In her reply Martina Viarengo reported that the interaction between school autonomy and accountability is insignificant in their analysis. She noted that although they have a rich dataset, limits still exist in terms of available controls. Viarengo acknowledged the argument that differences in the way that tracking is implemented may exist across countries. On the private–public divide, she told the panel that this matter is addressed in robustness checks not currently reported in the paper.

APPENDIX

Variables

The list below describes the variables used in this study that have been extracted from the PISA database (<http://pisa2009.acer.edu.au>). A thorough description is provided in the School, Student Questionnaires, the related Codebooks and PISA 2009 Technical Report (OECD, 2010d, 2012a). The variable related to tracking has been extracted from various sources and is available from the authors.

Tracking

- *School Tracking Policy* is a dummy variable =1 if age at first selection is below 16.
- *Tracking Index* is based on the actual age at first selection. The index varies between 10 and 18. Ages have been reverse coded.

Other sorting mechanisms

- *Private School* is a dummy variable = 1 if the school is private.
- *Private Funding* is the share of funding provided by private sources.
- *Selective Admissions* is a dummy variable = 1 if student's record of academic performance (including placement tests) is always considered when students are admitted to the school.
- *Selective Residence* is a dummy variable = 1 if residence in a particular area is always considered when students are admitted to the school.

Policies

School autonomy

- The *Index of the Relative Level of Responsibility of School Staff Relating to Curriculum and Assessment* was computed from four items measuring the school principal's report concerning who had responsibility for curriculum and assessment. The index was standardized to having an OECD mean of zero and a standard deviation of one.
- The *Index of the Relative Level of Responsibility of School Staff in Allocating Resources* was derived from six items measuring the school principals' report on who has considerable responsibility for tasks regarding school management of resource allocation. The index was standardized to having an OECD mean of zero and a standard deviation of one.

Resources

- The *Index on the School's Educational Resources* was computed on the basis of seven items measuring the school principal's perceptions of potential factors hindering instruction at school. All items were inverted for IRT scaling and positive WLE scores indicate better quality of educational resources.
- The *Index of Extra-Curricular Activities Offered by School* is based on school principals' reports of what extra-curricular activities occur at their school. Responses to the items were coded such that positive WLE scores indicate higher levels of extra-curricular school activities.
- The *Student-Teacher Ratio* is enrolment by the total number of teachers.

Teacher characteristics

- The *Proportion of Fully Certified Teachers* was computed by dividing the number of fully certified teachers by the total number of teachers.
- The *Proportion of Teachers with a University Education* is the number of teachers with ISCED 5A qualification divided by the total number of teachers.

Selectivity

- The *Index of Ability Grouping between Classes* was derived from the two items by assigning schools to three categories: (1) schools with no ability grouping for any subjects, (2) schools with one of these forms of ability grouping between classes for some subjects and (3) schools with one of these forms of ability grouping for all subjects.
- The *Index of Academic School Selectivity* was computed by assigning schools to three different categories: (1) schools where neither of these two factors is considered for student admittance, (2) schools considering at least one of these two factors, (3) schools where at least one of these two factors is a prerequisite for student admittance.

Accountability

- *Achievement data are posted publicly* is a dummy variable = 1 if school principals said achievement data was posted publicly during the past year.
- *Achievement data are used in evaluation of the principal's performance* is a dummy variable = 1 if school principals said achievement data was used to monitor the principal's performance during the past year.
- *Achievement data are used in evaluation of teachers' performance* is a dummy variable = 1 if school principals said achievement data was used to monitor teachers' performance during the past year.
- *Teacher peer-review* is a dummy variable = 1 if school principals said teacher peer review was used to monitor teachers during the past year.
- *External observers* is a dummy variable = 1 if school principals said external observers were used to monitor teachers during the past year.

Practices

Approaches to learning

- *Index of Memorization*. Positive WLE scores on these indices indicate higher importance attached to the given reading strategy. Four items measuring the construct of memorization were included in the PISA 2009 main study.
- *Index of Elaboration*. Positive WLE scores on these indices indicate higher importance attached to the given reading strategy. Four items measuring the construct of elaboration were included in the PISA 2009 main study.
- *Index of Control Learning Strategies*. Positive WLE scores on these indices indicate higher importance attached to the given reading strategy. Five items measuring the construct of control learning strategies were included in the PISA 2009 main study.

Staff attitude

- *Teachers' Low Expectations* as detrimental to the learning of students is a dummy variable = 1 if school principals said it was to some extent or a lot.
- *Staff Resisting Change* as detrimental to the learning of students is a dummy variable = 1 if school principals said to some extent or a lot.

Family background

- *Years of Schooling* is the highest level of education of either parent according to the ISCED classification.
- *Books at Home* provides the share of individuals according to the number of books they report having in their home.
- The *Index of Economic, Social and Cultural Status* is constructed by using three sub-components: the higher parental occupation index, the higher parental education index and the index of home possessions.

Parental involvement

- The *Index of the Early Activities with the Child when the Child attended the First Year of ISCED 1* is related to parental support of child's reading literacy and based on the information provided by parents. Positive WLE scores indicate greater parental support of child's reading literacy at the beginning of ISCED 1.
- The *Index of the Current Activities with the Child* is related to parental support in home activities and is based on the information provided by parents. Positive WLE scores indicate greater parental support of child's reading literacy.
- The *Index of Parental Involvement* is related to parents' participation in school-related activities and is based on the information provided by parents. Positive WLE scores indicate greater parental involvement in their child's school.
- The *Index of Resources Available to the Child* is related to resources available to the child at home and is based on the information provided by parents. Positive WLE scores indicate greater availability of reading resources at home.

Table A1. Descriptive statistics of the control variables, PISA 2009

| Country group | Students | | | | | Schools | | | | | | |
|-------------------|------------|--------------|--------------------|---------------------------------------|---------------------------------|---------------------------------|--|--|--------------------------|--------------------------|-----------------------|-----------------------|
| | Female (%) | Age | Grade ^a | Immigrant background (%) ^b | Mother immigrant background (%) | Father immigrant background (%) | Language of test spoken at home (%) ^c | Live with two parents (%) ^d | School size ^e | Village (%) ^f | Town (%) ^g | City (%) ^h |
| All | 50.39 | 15.78 (0.29) | -0.06 (0.72) | 3.80 | 9.50 | 9.58 | 85.02 | 75.46 | 967.82 (812.80) | 12.69 | 46.80 | 40.50 |
| OECD | 49.34 | 15.79 (0.29) | -0.07 (0.60) | 6.41 | 17.09 | 17.32 | 90.18 | 78.45 | 995.21 (719.58) | 9.19 | 53.60 | 37.21 |
| EU15+ | 49.90 | 15.79 (0.29) | -0.16 (0.62) | 6.85 | 16.20 | 16.45 | 90.04 | 81.88 | 761.55 (464.77) | 5.63 | 66.53 | 27.83 |
| Asian | 48.29 | 15.75 (0.29) | -0.10 (0.34) | 1.54 | 3.35 | 2.61 | 95.49 | 84.40 | 1159.59 (922.74) | 0.26 | 23.77 | 75.97 |
| Other high income | 50.29 | 15.80 (0.29) | 0.15 (0.65) | 7.29 | 17.53 | 17.97 | 90.17 | 71.78 | 559.60 (490.22) | 19.47 | 35.60 | 44.93 |
| Middle income | 51.90 | 15.77 (0.29) | -0.07 (0.87) | 1.22 | 2.26 | 2.36 | 76.92 | 70.77 | 959.62 (880.66) | 17.41 | 47.60 | 35.00 |

Notes: Mean value weighted by sampling probabilities; standard deviation (only for discrete variables).

^aGrade is compared to modal grade in country (classified as 0).

^bStudents foreign-born, that is born in a country different from country of test.

^cLanguage of test spoken at home most of the time.

^dLive with two (natural or otherwise) parents.

^eTotal school enrolment.

^fCommunity size where the school is located: a village, hamlet or rural area (fewer than 3,000 people).

^gCommunity size where the school is located: either a small town (from 3,000 to about 15,000 people) or a town (from 15,000 to about 100,000 people).

^hCommunity size where the school is located: either a city (from 100,000 to 1 million people) or a large city (with over 1 million people).

Table A2. School policies and practices, PISA 2009

| A. Policies | | | | | | | | | |
|-------------------|--|--|---|--|--|---------------------------------------|---|------------------------|--------------------|
| Country | School autonomy | | Resources | | | Selectivity | | | |
| | Responsibility over curriculum and assessment ^a | Responsibility over resource allocation ^a | Teacher-student ratio ⁱ | Quality of educational resources ^a | Extra-curricular activities ⁱ | Academic selectivity ^b (%) | Ability grouping between classes ^c (%) | | |
| All | -0.18 (0.98) | -0.12 (0.93) | 17.77 (11.92) | -0.28 (1.20) | 0.36 (1.03) | 41.45 | 18.97 | | |
| OECD | -0.14 (0.96) | 0.05 (0.98) | 14.82 (5.98) | 0.14 (1.08) | 0.54 (0.92) | 33.46 | 11.36 | | |
| EU15+ | 0.04 (0.96) | -0.10 (1.00) | 12.81 (4.79) | 0.10 (0.93) | 0.21 (0.89) | 36.43 | 12.16 | | |
| Asian | 0.81 (0.83) | -0.17 (0.98) | 14.62 (5.22) | 0.36 (1.02) | 0.52 (0.92) | 71.82 | 8.42 | | |
| Other high income | -0.37 (0.80) | -0.09 (0.70) | 11.86 (4.51) | -0.57 (0.96) | 0.70 (0.88) | 26.52 | 36.23 | | |
| Middle income | -0.41 (0.92) | -0.27 (0.88) | 22.24 (15.94) | -0.76 (1.17) | 0.10 (1.12) | 43.83 | 25.97 | | |
| Teachers | | | | | | | | | |
| Country | Share of qualified teachers | | Accountability | | Achievement data | | Teacher peer-review | | External observers |
| | Share of certified teachers | Achievement data posted publicly (%) | Achievement data used to evaluate principal (%) | Achievement data used to evaluate teachers (%) | Teacher peer-review (%) | Teacher peer-review (%) | Teacher peer-review (%) | External observers (%) | |
| All | 82.26 | 75.48 | 42.59 | 50.33 | 60.49 | 66.32 | 66.32 | 44.97 | |
| OECD | 83.67 | 82.81 | 53.03 | 44.29 | 42.92 | 50.00 | 50.00 | 34.47 | |
| EU15+ | 69.22 | 87.68 | 26.44 | 26.65 | 33.06 | 42.11 | 42.11 | 24.38 | |
| Asian | 96.05 | 96.75 | 16.38 | 18.07 | 34.45 | 60.21 | 60.21 | 36.41 | |
| Other high income | 86.90 | 94.79 | 70.42 | 76.79 | 93.84 | 93.19 | 93.19 | 68.86 | |
| Middle income | 77.07 | 60.63 | 33.99 | 59.47 | 78.59 | 79.68 | 79.68 | 53.40 | |

Table A2. (Continued)

| B. Practices | Approaches to learning | | Staff attitude | |
|-------------------|---------------------------|--------------------------|--|--|
| | Memorization ^a | Elaboration ^a | Control of learning strategies ^a | Staff resisting change ^b |
| Country | | | Teachers' low expectations ^b (% yes) | Staff resisting change ^b (% yes) |
| All | 0.09 (0.97) | 0.10 (0.98) | -0.04 (0.97) | 27.98 |
| OECD | 0.02 (1.00) | -0.01 (1.02) | 0.05 (1.02) | 31.01 |
| EU15+ | 0.02 (0.94) | -0.04 (0.96) | 0.07 (0.96) | 24.41 |
| Asian | -0.34 (1.05) | -0.31 (1.09) | -0.41 (1.02) | 36.18 |
| Other high income | 0.22 (0.85) | 0.09 (0.93) | -0.11 (0.86) | 33.24 |
| Middle income | 0.24 (0.89) | 0.30 (0.88) | -0.03 (0.90) | 25.06 |

Note: Sampling probability weights have been used in order to provide descriptive statistics representative for each country. Policy and Practice variables are described in the Appendix.

^aMean of the index and standard deviation (in parentheses).

^bAt least one factor is considered for school admittance (see variable description).

^cFor all subjects.

Sampling probability weights have been used in order to provide descriptive statistics representative for each country. Policy and Practice variables are described in the Appendix.

^aMean of the index and standard deviation (in parentheses).

^bPerceived by the school principal as detrimental to the learning of students.

REFERENCES

- Algan, Y., P. Cahuc and A. Shleifer (2013). 'Teaching practices and social capital', *American Economic Journal: Applied Economics*, 5(3), 189–210.
- Altinok, N. and G. Kingdon (2012). 'New evidence on class size effects: a pupil fixed effects approach', *Oxford Bulletin of Economics and Statistics*, 74(2), 203–34.
- Altonji, J. G., T. E. Elder and C. R. Taber (2005). 'Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools', *Journal of Political Economy*, 113 (1), 151–84.
- Ammermueller, A., H. Heijke and L. Woessmann (2005). 'Schooling quality in Eastern Europe: educational production during transition', *Economics of Education Review*, 24(5), 579–99.
- Angrist, J., E. Bettinger, E. Bloom, E. King and M. Kremer (2002). 'Vouchers for private schooling in Colombia: evidence from a randomized natural experiment', *American Economic Review*, 92(5), 1535–58.
- Bertola, G., and D. Checchi (2013). 'Who chooses which private education? Theory and international evidence', *Labour*, 27(3), 249–71.
- Betts, J. R. (2011). 'The economics of tracking in education', in E.A. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education*, Vol. 3, Chapter 7, North-Holland, Amsterdam.
- Betts, J. R. and J. L. Shkolnik (2000). 'The effects of ability grouping on student math achievement and resource allocation in secondary schools', *Economics of Education Review*, 19, 1–15.
- Bietenbeck, J. C. (2011). 'Teaching practices and student achievement: evidence from TIMSS', Master's thesis CEMFI No. 1104.
- Bishop, J. H. (1997). 'The effect of national standards and curriculum-based exams on achievement', *American Economic Review*, 87, 260–64.
- (2006). 'Drinking from the fountain of knowledge: student incentive to study and learn – externalities, information problems, and peer pressure', in E. A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, pp. 909–44, North Holland, Amsterdam.
- Bjorklund, A. and K. G. Salvanes (2010). 'Education and family background: mechanisms and policies', IZA Discussion Paper No. 5002.
- (2011). 'Education and family background: mechanisms and policies', in E. A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Chapter 3, North Holland, Amsterdam.
- Bourdieu, P. (1986). 'The forms of capital' (trans. R. Nice), in J. C. Richardson (ed.), *Handbook of Theory and Research for the Sociology of Education*, pp. 241–58, Greenwood Press, New York.
- Braga, M., D. Checchi and E. Meschi (2013). 'Educational policies in a long-run perspective', *Economic Policy*, 28(73), 45–100.
- Bratti, M. and D. Checchi (2013). 'Re-testing PISA students one year later: on school value added estimation using OECD-PISA', IZA Discussion Paper No. 7722.
- Bratti, M., D. Checchi and A. Filippin (2011). 'Should you compete or cooperate with your schoolmates?', *Education Economics*, 19(3), 275–89.
- Brunello, G. and D. Checchi (2007). 'Does school tracking affect equality of opportunity? New international evidence', *Economic Policy*, 22(52), 781–861.
- Brunello, G. and M. Schlotter (2011). 'Non cognitive skills and personality traits: labour market relevance and their development in education and training systems', IZA Discussion Paper No. 5743.
- Chingos, M. M. and G. J. Whitehurst (2011). 'Class size: what research says and what it means for state policy', www.brookings.edu/research/papers/2011/05/11-class-size-whitehurst-chingos
- Coleman, J. S. (1988). 'Social capital in the creation of human capital', *American Journal of Sociology*, 94 (Supplement), 95–120.
- Costa, P. T. Jr and McRae R. R. (1992). 'Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual', Psychological Assessment Resources, Inc., Odessa, Florida.
- Davis-Kean, P. E. (2005). 'The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment', *Journal of Family Psychology*, 19, 294–304.

- Desforges, C. and A. Abouchaar (2003). 'The impact of parental involvement, parental support and family education on pupil achievements and adjustment: a literature review', Department for Education and Skills, Nottingham.
- DiMaggio, P. (1982). 'Cultural capital and school success: the impact of status culture participation on the grades of U.S. High School Students', *American Sociological Review*, 47, 189–201.
- Dolton, P. (2006). 'Teacher supply', in E.A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Chapter 19, North-Holland, Amsterdam.
- Dronkers, J. (1992). 'Parents, love, and money: the relations between parental class, cognitive skill, educational attainment, occupation, marriage, and family income', *International Perspectives on Education and Society*, 2, 277–93.
- Dustmann, C., T. Frattini and G. Lanzara (2012). 'Educational achievement of second-generation immigrants: an international comparison', *Economic Policy*, 27 (69), 143–85.
- Epplé, D. and R. Romano (1998). 'Competition between private and public schools, vouchers, and peer group effects', *American Economic Review*, 88(1), 33–62.
- Figlio, D. N. and M. E. Page (2002). 'School choice and the distributional effects of ability tracking: does separation increase inequality?', *Journal of Urban Economics*, 51(3), 497–514.
- Freeman, R. B., S. Machin and M. Viarengo (2011). 'Inequality of educational outcomes: international evidence from PISA', *Regional and Sectoral Economic Studies*, 11(3), 5–20.
- Fuchs, T. and L. Woessmann (2007). 'What accounts for international differences in student performance? A re-examination using PISA data', *Empirical Economics*, 32(2–3), 433–62.
- George, R. and D. Kaplan (1998). 'A structural model of parent and teacher influences on science attitudes of eighth graders: evidence from NELS: 88', *Science Education*, 82, 93–109.
- Greene, J. P. and M. Winters (2005). 'Public high school graduation and college readiness: 1991–2002', Manhattan Institute for Policy Research, New York.
- Hanushek, E. A. (1979). 'Conceptual and empirical issues in the estimation of educational production functions', *Journal of Human Resources*, 14(3), 351–88.
- (1986). 'The economics of schooling: production and efficiency in public schools', *Journal of Economic Literature*, 24(3), 1141–77.
- (2006). 'School resources', in E. A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, 865–908, North-Holland, Amsterdam.
- Hanushek, E. A. and M. F. Raymond (2003). 'Lessons about the design of state accountability systems', in P. E. Peterson and M. R. West (eds.), *No Child Left Behind? The Politics and Practice of Accountability*, 127–51, Brookings Institute, Washington, DC.
- Hanushek, E. A. and S. G. Rivkin (2006). 'Teacher quality', in E. A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, 1, 1050–78, North Holland, Amsterdam.
- Hanushek, E. A. and L. Woessmann (2006). 'Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries', *Economic Journal*, 116 (510), C63–C76.
- (2011a). 'The economics of international differences in educational achievement', in E. A. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook in Economics*, Vol. 3, 89–200, North-Holland, The Netherlands.
- (2011b). 'Educational outcomes', *Economic Policy*, July, 427–91.
- Hanushek, E. A., S. Link and L. Woessmann (2013). 'Does school autonomy make sense everywhere? *Journal of Development Economics*, 104, 212–32.
- Holmlund, H., S. McNally and M. Viarengo (2010). 'Does money matter for schools?', *Economics of Education Review*, 29, 1154–64.
- Hoover-Dempsey, K. V. and H. M. Sandler (1997). 'Why do parents become involved in their children's education?' *Review of Educational Research*, 67(1), 3–42.
- Jackson, C. K. (2010). 'Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago', *Economic Journal*, 120(549), 1399–429.
- Jenkins, S. P., J. Micklewright and S. V. Schnepf (2008). 'Social segregation in secondary schools: how does England compare with other countries?', *Oxford Review of Education*, 34(1), 21–38.
- Jerrim, J. and J. Micklewright (2012). 'Socioeconomic gradients in children's cognitive skills: are cross-country comparisons robust to who reports family background?' DoQSS – Institute of Education – Working Paper No. 12-06.

- Jurges, H., K. Schneider and F. Buchel (2005). 'The effect of central exit examinations on student achievement: quasi-experimental evidence from TIMSS Germany', *Journal of the European Economic Association*, 3(5), 1134–55.
- Krueger, A. (1999). 'Experimental estimates of education production functions', *Quarterly Journal of Economics*, 114, 497–532.
- Lankford, L., S. Loeb and J. Wyckoff (2002). 'Teacher sorting and the plight of urban schools: a descriptive analysis', *Educational Evaluation and Policy Analysis*, 24(1), 37–62.
- Lavy, V. (2011). 'What makes an effective teacher? Quasi-experimental evidence', NBER Working Paper No. 16885.
- Machin, S. and S. McNally (2012). 'The evaluation of English education policies', *National Institute Economic Review*, 219(1), R15–R25.
- Nechyba, T. J. (2011). 'Income and peer quality sorting in public and private schools', in E. A. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education*, Vol. 3, Chapter 22, North-Holland, Amsterdam.
- Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality*, 2nd edn, Yale University Press, New Haven.
- OECD (2005). *Education at a Glance 2005*, OECD Publishing, Paris.
- (2010a). *PISA 2009 at a Glance*, OECD Publishing, Paris.
- (2010b). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)*, OECD Publishing, Paris.
- (2010c). *PISA 2009 Results: Learning Trends: Changes in Student Performance since 2000 (Volume V)*, OECD Publishing, Paris.
- (2010d). *PISA 2009 Results: What Makes a School Successful? – Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.
- (2012a). *PISA 2009 Technical Report*, OECD Publishing, Paris.
- (2012b). 'Parental Involvement in Selected PISA Countries and Economies', OECD Education Working Paper No. 73 prepared by F. Borgonovi and G. Montt.
- Ou, D. (2010). 'To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam', *Economics of Education Review*, 29(2), 171–86.
- Peterson, P. E. and F. M. Hess (2006). 'Keeping an eye on state standards', *Education Next*, 6(3), 28–29.
- (2008). 'Few states set world-class standards: In fact, most render the notion of proficiency meaningless', *Education Next*, 8(3), 70–3.
- Peterson, P. E. and M. Viarengo (2011). 'Social capital and school impacts on the non-cognitive skills of early adolescents: evidence from a national longitudinal survey', in M. Berends, M. Cannata and E. B. Goldring (eds.), *School Choice and School Improvement*, Chapter 4, Cambridge, MA, Harvard University Press.
- Peterson, P. E. and L. Woessmann (2007). 'Introduction: schools and the equal opportunity problem', in L. Woessmann and P. E. Peterson (eds.), *Schools and the Equal Opportunity Problem*, 3–27, MIT Press, Cambridge, MA.
- Peterson, P. E., L. Woessmann, E. A. Hanushek and C. X. Lastra-Anadón (2011). 'Globally challenged: are U. S. students ready to compete? The latest on each state's international standing in math and reading', Harvard's PEPG Report No. 11-03.
- Pop-Eleches, C. and M. Urquiola (2013). 'Going to a better school: effects and behavioral responses', *American Economic Review*, 103(4), 1289–324.
- Schoon, I. and S. Parsons (2002). 'Competence in the face of adversity: the influence of early family environment and long-term consequences', *Children and Society*, 16, 260–72.
- Schuetz, G., H. W. Ursprung and L. Woessmann (2008). 'Education policy and equality of opportunity', *Kyklos*, 61(2), 279–308.
- Schwerdt, G. and A. C. Wuppermann (2011). 'Is traditional teaching really all that bad? A within-student between-subject approach', *Economics of Education Review*, 30, 365–79.
- Singh, K., P. G. Bickley, T. Z. Keith, P. B. Keith, P. Trivette and E. Anderson (1995). 'The effects of four components of parental involvement on eighth-grade student achievement: structural analysis of NELS-88 data', *School Psychology Review*, 24(2), 299–317.
- Sprietsma, M. (2008). 'Regional school choice and school selectivity: how do they relate to student performance? Evidence from PISA 2003', *European Journal of Comparative Economics*, 5(2), 133–56.

- West, M. R. and L. Woessmann (2006). 'Which school systems sort weaker students into smaller classes? International evidence', *European Journal of Political Economy*, 22(4), 944–68.
- Woessmann, L. (2003a). 'Schooling resources, educational institutions, and student performance: the international evidence', *Oxford Bulletin of Economics and Statistics*, 65(2), 117–70.
- (2003b). 'Central exit exams and student achievement: international evidence', in P. E. Peterson and M. R. West (eds.), *No Child Left Behind? The Politics and Practice of School Accountability*, 292–323, Brookings Institution Press, Washington, DC.
- (2005a). 'The effect heterogeneity of central exams: evidence from TIMSS, TIMSS-Repeat and PISA', *Education Economics*, 13(2), 143–69.
- (2005b). 'Educational production in East Asia: the impact of family background and schooling policies on student performance', *German Economic Review*, 6(3), 331–53; Kiel Working Paper No. 1152.
- (2008). 'How equal are educational opportunities? Family background and student achievement in Europe and the United States', *Zeitschrift für Betriebswirtschaft*, 78(1), 45–70.
- (2009). 'Public–private partnerships and student achievement: a cross-country analysis', in R. Chakrabarti and P. E. Peterson (eds.), *School Choice International: Exploring Public-Private Partnerships*, 13–45, MIT Press, Cambridge, MA.
- (2010). 'Families, schools, and primary-school learning: evidence for Argentina and Colombia in an international perspective', *Applied Economics*, 42(21), 2645–65.
- Woessmann, L., E. Ludemann, G. Schuetz and M. West (2007). 'School accountability, autonomy, choice, and the level of student achievement: international evidence from PISA 2003', OECD Education Working Papers No. 13.
- Woessmann, L., E. Ludemann, G. Schuetz and M. West (2009). *School Accountability, Autonomy, and Choice around the World*, Edward Elgar, Cheltenham.